



# Essays on Econometrics and Decision Theory

## Citation

Montiel Olea, Jose Luis. 2013. Essays on Econometrics and Decision Theory. Doctoral dissertation, Harvard University.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11107808>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

© 2013 José Luis Montiel Olea

All rights reserved.

*Dissertation Advisors:*

**Professor Gary Chamberlain**  
**Professor Tomasz Strzalecki**

*Author:*

**José Luis Montiel Olea**

## **Essays on Econometrics and Decision Theory**

### **Abstract**

This dissertation presents three essays. The first essay, coauthored with Tomasz Strzalecki, is a classical exercise in axiomatic decision theory. We propose a simple and novel axiomatization of *quasi-hyperbolic discounting*, a tractable model of present bias preferences that has found many applications in economics. Our axiomatization imposes consistency restrictions directly on the intertemporal tradeoffs faced by the decision maker, without relying on auxiliary calibration devices such as lotteries. Such axiomatization is useful for experimental work since it renders the short-run and long-run discount factor elicitation independent of assumptions on the decision maker's utility function.

The second essay, coauthored with Carolin Pflueger, belongs to the field of econometric theory. We develop a test for weak identification in the context of linear instrumental variables regression. The central feature of our test is its robustness to heteroskedasticity, autocorrelation, and clustering. We define identification to be weak when the Two-Stage Least Squares (TSLS) or the Limited Information Maximum Likelihood (LIML) Nagar bias is large relative to a benchmark. To test the null hypothesis of weak identification we propose a scaled non-robust first stage F statistic: *the effective F*. The test rejects for large values of the effective F. The critical values depend on an estimate of the covariance matrix of the OLS reduced form regression coefficients

and on the covariance matrix of the reduced form errors.

The third essay—the main chapter of this dissertation—belongs to the intersection of econometric theory and statistical decision theory. I present a new class of tests for hypothesis testing problems with a special feature: a *boundary-sufficient statistic*. The new tests minimize a weighted sum of the average rates of Type I and Type II error (*average risk*), while controlling the conditional rejection probability on the boundary of the null hypothesis; in this sense they are *efficient conditionally similar on the boundary* (ECS). The ECS tests emerge from an axiomatic approach: they essentially characterize *admissibility*—an important finite-sample optimality property—and *similarity on the boundary* in the class of all tests, provided the boundary-sufficient statistic is *boundedly complete*.



# Contents

Abstract . . . . .	iii
Acknowledgments . . . . .	x
<b>1 A Simple Axiomatization of Quasi-Hyperbolic Discounting</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Preliminaries . . . . .	5
1.3 Axiomatic Characterization . . . . .	6
1.3.1 Compensation Axiom . . . . .	9
1.3.2 Alternate Approaches . . . . .	11
1.4 Experimental Design . . . . .	14
1.4.1 Alternate Design . . . . .	16
1.5 Related Literature . . . . .	17
<b>2 A Robust Test for Weak Instruments</b>	<b>19</b>
2.1 Introduction . . . . .	19
2.2 Model and Summary of Testing Procedure . . . . .	23
2.2.1 Model and Assumptions . . . . .	23
2.2.2 Implementing the Testing Procedure . . . . .	25
2.3 Asymptotic Distributions and an Example . . . . .	29
2.3.1 Illustrative Example . . . . .	29
2.3.2 Asymptotic Distributions . . . . .	31
2.4 Testing the Null Hypothesis of Weak Instruments . . . . .	33
2.5 Computation of Critical Values . . . . .	37
2.6 Empirical Application: Elasticity of Intertemporal Substitution . . . . .	40
2.7 Conclusion . . . . .	44
<b>3 Efficient Conditionally Similar-on-the-Boundary Tests</b>	<b>45</b>
3.1 Introduction . . . . .	45
3.2 Basic Definitions and Assumptions . . . . .	50

3.2.1	Basic Elements of a Parametric Testing Problem . . . . .	51
3.2.2	Tests, Type I/Type II Error and Risk Function . . . . .	54
3.3	Testing problems with a boundary-sufficient statistic . . . . .	56
3.4	Main Results . . . . .	58
3.4.1	ECS Tests . . . . .	59
3.4.2	Theorem 1 . . . . .	61
3.4.3	Theorem 2 . . . . .	65
3.5	Examples . . . . .	69
3.5.1	Linear Instrumental Variables Regression (IV) . . . . .	70
3.5.2	Weakly Identified GMM Models . . . . .	82
3.5.3	Dynamic Effects in Structural VARs . . . . .	91
3.6	Summary of the main results and Conclusion . . . . .	97
<b>References</b>		<b>102</b>
<b>Appendix A Proofs</b>		<b>110</b>
A.1	Proofs for Chapter 1 . . . . .	110
A.1.1	Proof of Theorem 1 . . . . .	110
A.1.2	Proof of Theorem 2 . . . . .	114
A.1.3	Proof of Theorem 3 . . . . .	115
A.1.4	Proof of Theorem 4 . . . . .	116
A.1.5	Proof of Theorem 5 . . . . .	119
A.2	Proofs for Chapter 2 . . . . .	123
A.2.1	Proof of Lemma 1 . . . . .	123
A.2.2	LIML Distribution in Illustrative Example . . . . .	124
A.2.3	Proof of Theorem 1 . . . . .	125
A.2.4	Proof of Lemma 2 . . . . .	128
A.3	Proofs for Chapter 3 . . . . .	131
A.3.1	Proof of Lemma 1 . . . . .	131
A.3.2	Proof of Theorem 1 . . . . .	135
A.3.3	Proof of Theorem 2 . . . . .	140
A.3.4	Example 1: Boundary Sufficiency for the IV model . . . . .	142
A.3.5	Proof of Result 1 . . . . .	144
A.3.6	Proof of Result 2 . . . . .	147
A.3.7	Proof of Result 2* . . . . .	155
A.3.8	Proof of Lemma 3 . . . . .	157

A.3.9	Proof of Result 3 . . . . .	159
A.3.10	Proof of Result 4 . . . . .	160
A.3.11	Proof of Result 5 . . . . .	161
A.3.12	Proof of Result 6 . . . . .	162
<b>Appendix B Supplementary Material</b>		<b>166</b>
B.1	Supplementary Material for Chapter 2 . . . . .	166
B.1.1	Numerical Accuracy of Patnaik Methodology . . . . .	166
B.1.2	Monte-Carlo Method . . . . .	168
B.1.3	Levels and Sizes of Patnaik and Monte Carlo Critical Values . .	169
B.1.4	Numerical Implementation of $B_{TSLS}(\mathbf{W}, \mathbf{\Omega})$ and $B_{LIML}(\mathbf{W}, \mathbf{\Omega})$ .	170
B.1.5	Comparing Robust Critical Values to Stock and Yogo(05) . . . .	171
B.1.6	Nagar Bias Approximates Asymptotic Bias . . . . .	174
B.1.7	Primitive Conditions for Independent Data . . . . .	176
B.1.8	Uniformity . . . . .	180
B.1.9	Uniformity problems with pretests . . . . .	181
B.1.10	Uniformity and Tests for Weak Instrumets . . . . .	182

## List of Tables

2.1	Critical Values . . . . .	27
2.2	Estimating the Elasticity of Intertemporal Substitution . . . . .	42
B.1	Numerical Accuracy of Patnaik Methodology . . . . .	167
B.2	TSLS and LIML critical values . . . . .	173

## List of Figures

2.1	TSLS and simplified 5% critical values . . . . .	28
3.1	5% Conditional Critical Region (AR,LM); $k = 2$ , $\lambda^2=1$ . . . . .	78
B.1	Rejection Probabilities for the Wald and T Tests ( $\alpha = 5\%$ ) . . . . .	185
B.2	Rejection Probabilities for the Wald and T Tests ( $\alpha = 5\%$ ), ctd. . . . .	186

## Acknowledgments

Para comenzar y terminar este trabajo ha sido imprescindible el apoyo de mi familia, mis amigos, y mis profesores. Escuché hace algunos años que el objetivo central de un agradecimiento público es arruinar la reputación del que agradece: la mala memoria, por un lado, tomará por víctimas a algunos nombres importantes; la mala pluma, por el otro, se encargará de que el reparto de créditos y alabanzas no estén a la altura de las circunstancias. Creo en la naturaleza perversa de esta sección, y no me queda más alternativa que ofrecer sinceras disculpas por lo que pueda olvidar y lo que no sea capaz de expresar.

Primero, quiero agradecerle a mis padres, desde el fondo de mi corazón, por todo lo que me han dado: su cariño, sus valores, su ejemplo, su tiempo. Les agradezco por hacer de mí y de mi hermano el centro de su mundo. Aunque creo que nunca se los he dicho, me parece un buen momento para que sepan que no los cambiaría por nada ni por nadie. Lo digo muy en serio. Fueron ustedes, con su disciplina y dedicación, los que han terminado esta tesis. Fueron ustedes, con su fortaleza de espíritu, los que hace años terminaron una licenciatura y aplicaron al doctorado. Me tomó algún tiempo, pero al final he entendido que fueron ustedes; siempre fueron ustedes. Fue tu lucha constante, mamá. Fueron tus terceros turnos, papá...

Y también fue mi hermano, a quien le agradezco por ser protagonista en muchas de los mejores memorias de mi vida: los partidos de fútbol y los floreros rotos en la sala de la Céspedes; las dominadas después de comer en el patio de la casa; el basquetbol en el tablero de Shaq; todas las tardes de televisión y caricaturas; nuestros partidos de ajedrez (¡los voluntarios y los comprados!), los de tenis, los de pin; los domingos de película en Santa Teresa.

He sido muy afortunado en ser parte de dos familias extensas y maravillosas: los

Olea Iturralde y los Montiel Ubaldo. Les agradezco a mis tíos, primos, y sobrinos por todos por todas las marcas que han dejado en mí, por todas las conversaciones (especialmente, aquellas con mi tío Angel), por las navidades, los años nuevos, y todos los buenos momentos. Un agradecimiento muy especial para mi tía Mary, quien se convirtió en el mecenas más generoso que pude tener. Te agradezco todo el apoyo que me brindaste antes y durante la universidad; además de toda la ayuda que diste a mis padres en momentos difíciles.

El segundo grupo de agradecimientos es para todos mis amigos: los de Cambridge, los del ITAM, los de Pachuca. Mi vida ha sido muy divertida y plena gracias a ustedes. *Yuhta Ishii, Michal Kolesar, Timothy McQuade*: ha sido un placer compartir techo por más de tres años. *Miriam Huntley (Midi) y Fede Villalpando*: mis últimos dos años en el doctorado no hubieran sido los mismos sin mi familia del barrio; gracias por contagiarme su alegría por la vida, gracias por su apoyo y gracias por su amistad. *Raluca Dragusanu, Eduardo Dávila, Nathaniel Hilger, Lauren Merrill, Joana Naritomi, Arash Nekoei, Francisco Queiro, Ashwin Rashtogi, Elvira Rocha, Heather Schofield, Lucia Tian, Ian Tomb*: gracias por los buenos momentos en “The Donut”. Jo: el paseo en bicicleta en la laguna y nuestra plática de esa tarde se quedaran conmigo para siempre (aún cuando en el olvido se depositen nuestras palabras). Laurencita: gracias por tu paciencia, tu calidez, tu persona.

*Georgy Egorov, Stefano Giglio, Mihai Manea, Eduardo Morales*: gracias por todo su trabajo duro, fuente única de motivación. Gracias a mis amigos brasileños: *Eduardo Azevedo, Leonardo Bursztyn, Susana Cordeiro-Guerra, Guilherme Lichand, Juliana Plaster, Fernando Roriz, Tiago Severo*. Dudu: gracias por todo tu apoyo y tus consejos durante mi búsqueda de trabajo; estoy convencido que tu guía hizo una diferencia importante. Lio: gracias por todas las salidas en Central. Susanita: gracias por ser la persona que eres y por todas las atenciones y cuidados en Rio cuando enfermé. A

mis amigos franceses: *Paul Bougarde, Charles-Henri Weymuller, Thibault Peyronel*: gracias por las fiesta y la buena música del gueto. Gracias a mis amigos argentinos y uruguayos: *Bernarda Dall’Aglio, Sebastian Di Tella, Horacio Larreguy, Mercedes Politi, Juan Passadore, Laurita Trucco, Nico Pérez Truglia, Marina Nuñez, Fernando Yu, Juan Xandri*. Plumi: tenías mucha razón: ¡mejoraste ni calidad de vida! Gracias a todos mis amigos de Littauer: *Steve Cicala, Tim Ganser, Frank Machado, Mikkel Plagborg-Møller, Daniel Pollmann, Guillaume Pouliot, Stan Veuger*. Gracias a mis dos excelentes compañeras de fiesta: *Mariluz Bort y Dalila Colucci*. Gracias a mi políglota favorita de Boston: mi queridísima *Victoria(na) Livingstone*.

A mis compañeros de cuarto y hermanos del ITAM, *Erick Fosado y Luis Esteban Islas*, gracias por la amistad sincera e incondicional a lo largo de estos años. A mis amigos de la universidad: *Marco Antonio Corral Almada, José Arturo Bulos, Ponchillo Cebreneros, Francisco Javier Chapa, Montse Fosado, Pablo Germán Hill de Titto, Mauricio Hernandez Cervantes, Luis Monroy, Arturo Quirarte Dayarse, Pablo Ramos, Jerome Martin Rozat y David Sanchez Zepeda*. Titto: aprendí mucho de ti. Gracias, por supuesto, a todos los amigos mexicanos con los que compartí inviernos y veranos en Cambridge: *Arturo Aguilar, Jaime Alatorre, Xavier Alvarado, Sofia Barrera, Lorenza De Icaza, Elisa De Anda, Guillermo García, Lorelen De Urioste, Alison Elias, Ricardo Enriquez, Valeria Espinosa, Mauricio Fernández Duque, Helena García, Viviana García, Rodrigo Garza, Ricardo Godínez, Claudio González, Maisu Huebe, Julián Leon, Guillermo Lerdo de Tejada, Arturo Limón, Aldo Mussachio, Odette Maciel, Karla Montaña, Claudia Nuñez, María Helena Ortega, Claudia Ramírez, Esteban Rossi-Hansberg, Alonso Sanchez, Sergio Silva, José Francisco Ursua, Manolito Victoria, Alberto Villalpando y Luis Zermeno*. Luisillo: por exagerado que te parezca, tu persona se ha convertido en un modelo a seguir. Te admiro y respeto mucho. Betón: Muchas gracias por las parrilladas en tu casa y todas los buenos ratos que pasamos



juntos. Rodri: que te puedo decir, ¡a darle gas! Aldo y Esteban: gracias por el ejemplo y la inspiración. A mi banda de Pachuca: *Carlos Moreno, Alex Kamichi, Cali, Carlo (Perale), Chris y Tashiro*: gracias por el jaque y la prepa. Una mención con mucho cariño para *Silvia Albrizio*: he aprendido que en cualquier día, por malo que sea, basta con acordarme de ti para recuperar la sonrisa.

El último agradecimiento—no menos importante—es para mis asesores de tesis y mis profesores. Ellos han contribuido a mi formación técnica y, más importante aún, a mi formación como persona.

El capítulo central de esta tesis surgió después de diversas visitas a la oficina de Gary Chamberlain. Pocos momentos en mi vida académica han sido tan placenteros como las horas que pasé en su oficina durante mi cuarto año en el doctorado. A él le agradezco por el tiempo que dedicó a revisar mi trabajo y por la seriedad con la que lo hizo. Su trabajo y su persona me han dado una idea muy clara del econometrista en que me quiero convertir.

Aprender Teoría de Decisión de Tomasz Strzalecki y escribir un artículo con él ha sido una experiencia muy enriquecedora. Sin temor a exagerar, puedo decirles que he sido testigo de la creatividad de una mente maravillosa. En el futuro, me gustaría alcanzar esa felicidad interna que le caracteriza; me gustaría, como él, desbordar y transmitir tanta energía positiva a las personas que me rodean. A Tomasz le agradezco el tomarme con seriedad como alumno y preocuparse siempre por mi desarrollo personal y profesional.

Matías Cattaneo ha jugado un papel fundamental en la etapa final de mi doctorado. Le agradezco la apuesta que hizo por mí y por mi trabajo en un momento en que las cosas no parecían ir por buen camino. Le agradezco también sus llamadas en diciembre, antes de la convención en San Diego, para motivarme y asegurarse de que yo estuviera entero para las entrevistas de trabajo.

La decisión de hacer investigación en Econometría la tomé después de mi clase de segundo año con James Stock. A Jim le agradezco por la introducción a la econometría teórica y a la literatura de identificación débil. A pesar que motivos laborales han impedido que él forme parte del comité de mi tesis doctoral, puedo decir con mucho orgullo que mi investigación le debe mucho a su pensamiento y sus enseñanzas.

Otros de mis profesores han tenido también un efecto profundo en mi persona. El libro máspreciado de mi doctorado [Ferguson (1967)] llegó a mis manos gracias a una sugerencia de Ulrich Müller. Sin saberlo, fue él quien despertó mi interés en la estadística matemática. A él le agradezco su atinada sugerencia y su maravillosa clase.

Tuve la fortuna de acompañar a Andrei Shleifer en algunas de sus corridas habituales por el río Charles. Algunas veces hablamos de economía, algunas otras de economistas, y en ocasiones más entretenidas platicábamos de la vida. Le agradezco por siempre animarme cuando lo necesitaba y por compartir sus experiencias de vida. Le agradezco por tomarse en serio su papel como formador.

José Córdoba, Andrei Gornberg, Diego Dominguez, Alejandro Hernandez, Ignacio Lobato, Lorenza Martínez, Tridib Sharma y Ricard Torres escribieron mis cartas para el doctorado y fueron ellos los que me prepararon para esta etapa. Les estaré siempre muy agradecido. Nacho: gracias por los correos, las llamadas, y las divertidas visitas a tu oficina. Alejandro: como te dije hace poco, esta tesis arrancó hace muchos años con tu proyecto educativo. Doc: para describir con precisión las cosas que aprendí de ti me haría falta un capítulo adicional para este trabajo; baste con decir que aquellos tiempos en que trabajé contigo fueron de gran estímulo intelectual. Lorenza: ¡espero que trabajemos juntos otra vez!

14 de Mayo de 2013. Cambridge, MA.

*Para Mitos, para Joaquís, para Nati y para mi sagaz, audaz, sabia y  
querida abuelita Mariquita (persona de hechos finos y palabras elocuentes)*

# Chapter 1

## A Simple Axiomatization of Quasi-Hyperbolic Discounting <sup>1</sup>

### 1.1 Introduction

Understanding how agents trade off costs and benefits that occur at different periods of time is a fundamental issue in economics. For many years, the leading paradigm used for the analysis of intertemporal choice has been the discounted utility model introduced by Samuelson (1937) and first axiomatized by Koopmans (1960).

As is well known, there are two main properties of this utility representation: *time separability* and *stationarity*. Time separability requires that the marginal rate of substitution between any two periods be independent of the consumption levels in other periods. This rules out intertemporal complementarity, habit formation, and related phenomena. Stationarity requires that a ranking of two consumption streams remains the same if both streams are delayed by one period.

The *present bias*—now a well documented phenomenon—is a failure of stationarity

---

<sup>1</sup>Co-authored with Tomasz Strzalecki

in which the marginal rate of substitution between consumption in periods 0 and 1, is smaller than the marginal rate of substitution between periods 1 and 2. For example, the following preference pattern is indicative of present bias.

$$(10, 2, 0, 0, \dots) \succ (8, 8, 0, 0, \dots) \quad (1.1a)$$

and

$$(0, 10, 2, 0, 0, \dots) \prec (0, 8, 8, 0, 0, \dots), \quad (1.1b)$$

where both symbols  $\succ$  and  $\prec$  refer to the preference over consumption streams expressed at the beginning of time before receiving any payoffs.

It is well known that the present bias may lead to violations of dynamic consistency when choices at later points in time are included in the model. This paper is concerned only with the “time zero” choices, therefore its results can be used in combination with any of the auxiliary assumptions (naivete, sophistication, partial sophistication, costly self-control, etc.) used to tie together choices made at different points in time.

The present bias is a very intuitive and widely observed phenomenon and many utility representations that allow for this feature have been studied in the literature. *Quasi-hyperbolic discounting*, proposed by Phelps and Pollak (1968), is a simple and tractable model of present bias preferences that has found many applications in economics.<sup>2</sup> Quasi-hyperbolic discounting evaluates a consumption stream  $(x_0, x_1, x_2, \dots)$  by

$$V(x_0, x_1, x_2, \dots) = u(x_0) + \beta \sum_{t=1}^{\infty} \delta^t u(x_t),$$

where  $u$  is the flow utility function,  $\delta \in (0, 1)$  is the long-run discount factor, and  $\beta \in (0, 1]$  is the short-run discount factor that captures the strength of the present bias;  $\beta = 1$  corresponds to the standard discounted utility model. The above equation

---

<sup>2</sup>See, e.g., Laibson (1997), Laibson (1998), DellaVigna and Malmendier (2004), Diamond and Köszegi (2003), Laibson *et al.* (2007).

implies that quasi-hyperbolic discounting retains the property of time-separability but violates stationarity. It does so, however, in a minimal way: stationarity is satisfied from period  $t = 1$  onward, a property called quasi-stationarity. Our axiomatization relies on these two properties, as well as a third property that is closely tied to the experimental measurement of the parameter  $\beta$ .

The customary method of measuring the strength of the present bias focuses directly on the tradeoff between consumption levels in periods 0 and 1, see, eg., Thaler (1981). The value of  $\beta$  can be revealed by varying consumption in period 1 to obtain indifference to a fixed level of consumption in time 0. However, this inference relies on parametric assumptions about the utility function  $u$  and is subject to many experimental confounds, see, e.g., McClure *et al.* (2007) and Noor (2009, 2011) among others. Hayashi (2003) employs a conceptually related method that uses probability mixtures to elicit the tradeoffs. However, his method relies on the expected utility assumption and couples together the agent's risk aversion and his intertemporal elasticity of substitution.

The method that our axiomatization is building on uses only two fixed consumption levels, but instead varies the time horizon.<sup>3</sup> In the quasi-hyperbolic discounting model the subjective distance between periods 0 and 1 (measured by  $\beta\delta$ ) is larger than the subjective distance between periods 1 and 2 (measured by  $\delta$ ), which is the reason behind the preference pattern (1a)–(1b). We uncover the parameter  $\beta$  by increasing this latter objective gap enough to make it subjectively equal to the former. The size of the gap needed is directly related to the value of the parameter  $\beta$ ; for example, if  $\beta = \delta$ , then the gap between periods 0 and 1 ( $\beta\delta$ ) is equal to the gap between periods

---

<sup>3</sup>A related but distinct method of standard sequences was used by Loewenstein and Prelec (1992) and Attema *et al.* (2010).

1 and 3 ( $\delta^2$ ). In this case, the following preference pattern obtains:

$$(10, 2, 0, 0, \dots) \succ (8, 8, 0, 0, \dots) \quad (1.2a)$$

if and only if

$$(0, 10, 0, 2, 0, 0, \dots) \succ (0, 8, 0, 8, 0, 0, \dots). \quad (1.2b)$$

We show how finding the right spacing of payoffs in time makes it possible to uncover the value of  $\beta$ . Different different sizes of the temporal gap correspond different to values of  $\beta$ . Since we are working in discrete time, for certain values of  $\beta$  there may not exist a corresponding gap that would make the decision maker indifferent. However, by relying on time separability we develop a more general notion of *compensation* that makes this indifference possible and helps us uncover  $\beta$ .

The crucial aspect of our measurement method is that for any given  $\beta$  the size of the gap (or the compensation more generally) is independent of the utility function  $u$ . Thus, the measurement of  $\beta$  is independent of the measurement of  $u$ . This makes it possible to study how impatience depends on the consumption good. By focusing directly on the intertemporal tradeoffs instead of relying on the utility function the same elicitation method can be used for any consumption good. Comparing the discount factors obtained in such elicitation will reveal the dependence of impatience on the good the consumption of which is being delayed. Measuring the dependence of impatience on the consumption good would enhance our understanding of the present bias and may be helpful in investigating the relationships between the  $\beta$ - $\delta$  model and other models, such as Gul and Pesendorfer (2001), which capture temptation without focusing explicitly on discounting.

The separation of the two aspects of preferences (discounting from utility) that is obtained in this paper is also important on conceptual grounds. The model has

two kinds of parameters: discount factors,  $\beta$  and  $\delta$ , which measure impatience, and the utility function,  $u$ , which measures the intertemporal elasticity of substitution. These two properties of preferences are conceptually and behaviorally distinct and the proposed axiomatization and calibration respect this separation and lead to the identification of the discount factors by measuring the exact types of behavior that these factors are responsible for. The parameter  $\beta$  is responsible for inflating the subjective distance between periods 0 and 1, as compared to the distance between 1 and 2. The preference patterns like (2a)–(2b) reflect precisely this property. This makes for a clear-cut measurement in comparison to the alternative methods which mix-in other phenomena, such as consumption smoothing attitudes.

The rest of the paper is organized as follows. Section 1.2 presents the primitives and basic notation. Section 1.3 presents the axioms and the representation theorems. Section 1.4 presents the proposed method of experimental measurement of the parameters inspired by our axiomatization. Section 1.5 discusses the related literature. Proofs and additional results are collected in the Appendix.

## 1.2 Preliminaries

Let  $C$  be the set of possible consumption levels, formally a connected and separable topological space. The set  $C$  could be monetary payoffs, but also any other divisible good, such as juice (McClure *et al.*, 2007), or level of noise (Casari and Dragone, 2010). Let  $\mathcal{T} := \{0, 1, 2, \dots\}$  be the set of time periods. *Consumption streams* are members of  $C^{\mathcal{T}}$ . A consumption stream  $x$  is *constant* if  $x = (c, c, \dots)$  for some  $c \in C$ . For any  $c \in C$  we slightly abuse the notation by denoting the corresponding constant stream by  $c$  as well. For any  $a, b, c \in C$  and  $x \in C^{\mathcal{T}}$  the streams  $ax$ ,  $abx$ , and  $abcx$  denote  $(a, x_0, x_1, \dots)$ ,  $(a, b, x_0, x_1, \dots)$ , and  $(a, b, c, x_0, x_1, \dots)$  respectively.



For any  $T$  and  $x, y$  define  $x_T y = (x_0, x_1, \dots, x_T, y_{T+1}, y_{T+2}, \dots)$ . A consumption stream  $x$  is *ultimately constant* if  $x = x_T c$  for some  $T$  and  $c \in C$ . For any  $T$  let  $X_T$  denote the set of ultimately constant streams of length  $T$ . Any  $X_T$  is homeomorphic to  $C^{T+1}$ . Consider a preference  $\succsim$  defined on a subset  $\mathcal{F}$  of  $C^T$  that contains all ultimately constant streams. This preference represents the choices that the decision maker makes at the beginning of time before any payoffs are realized. We focus on preferences that have a quasi-hyperbolic discounting representation over the set of streams with finite discounted utility.

**Definition.** A preference  $\succsim$  on  $\mathcal{F}$  has a *quasi-hyperbolic discounting* representation if and only if there exists a nonconstant and continuous function  $u : C \rightarrow \mathbb{R}$  and parameters  $\beta \in (0, 1]$  and  $\delta \in (0, 1)$  such that  $\succsim$  is represented by the mapping

$$x \mapsto u(x_0) + \beta \sum_{t=1}^{\infty} \delta^t u(x_t).$$

As mentioned before, the parameter  $\beta$  can be thought of as a measure of the present bias. The parameter  $\beta$  represents the size of the subjective distance between periods 0 and 1. As we will see, this parameter has a clear behavioral interpretation in our axiom system and it will become explicit in what sense  $\beta$  is capturing the subjective distance between periods 0 and 1.

### 1.3 Axiomatic Characterization

Our axiomatic characterization involves two steps. First, by modifying the classic axiomatizations of the discounted utility model, we obtain a representation of the form:

$$x \mapsto u(x_0) + \sum_{t=1}^{\infty} \delta^t v(x_t) \tag{1.3}$$

for some nonconstant and continuous  $u, v : X \rightarrow \mathbb{R}$  and  $0 < \delta < 1$ . Second, we impose our main axiom to conclude that  $v(c) = \beta u(c)$  for some  $\beta \in (0, 1]$ .

Our axiomatization of the representation (1.3) builds on the classic work of Koopmans (1960, 1972), recently extended by Bleichrodt *et al.* (2008). The first axiom is standard.

**Axiom 1** (Weak Order).  $\succsim$  is complete and transitive.

The second axiom, sensitivity, guarantees that preferences are sensitive to payoffs in periods  $t = 0$  and  $t = 1$  (sensitivity to payoffs in subsequent periods follows from the quasi-stationarity axiom, to be discussed later). Sensitivity is a very natural requirement, to be expected of any class of preferences in the environment we are studying.

**Axiom 2** (Sensitivity). There exist  $e, c, c' \in C$  and  $x \in \mathcal{F}$  such that  $cx \succ c'x$  and  $ecx \succ ec'x$ .

The third axiom, initial separability, involves conditions that ensure the separability of preferences across time. (These conditions are imposed only on the few initial time periods, but extend beyond them as a consequence of the quasi-stationarity axiom.) Time separability is a necessary consequence of any additive representation of preferences and is not specific to quasi-hyperbolic discounting.

**Axiom 3** (Initial Separability). For all  $a, b, c, d, e, e' \in C$  and all  $z, z' \in \mathcal{F}$  we have

- (a)  $abz \succ cdz$  if and only if  $abz' \succ cdz'$ ,
- (b)  $eabz \succ ecdz$  if and only if  $eabz' \succ ecdz'$ ,
- (c)  $ex \succ ey$  if and only if  $e'x \succ e'y$ .

The standard geometric discounting preferences satisfy a requirement of *stationarity*, which says that the tradeoffs made at different points in time are resolved in the same way. Formally, stationarity means that  $cx \succ cy$  if and only if  $x \succ y$  for

any consumption level  $c \in C$  and streams  $x, y \in \mathcal{F}$ . However, as discussed in the introduction, the requirement of stationarity is not satisfied by quasi-hyperbolic discounting preferences; in fact, it is the violation of stationarity, that is often taken to be synonymous with quasi-hyperbolic discounting. Nevertheless, quasi-hyperbolic discounting preferences possess strong stationarity-like properties, since the preferences starting from period 1 onwards are geometric discounting.

**Axiom 4** (Quasistationarity). For all  $e, c \in C$  and all  $x, y \in \mathcal{F}$ ,  $ecx \succ ecy$  if and only if  $ex \succ ey$ .

The last three axioms, introduced by Bleichrodt *et al.* (2008), are used instead of stronger infinite dimensional continuity requirements. They are of technical nature, as are all continuity-like requirements. However, constant-equivalence and tail-continuity have simple interpretations in terms of choice behavior.

**Axiom 5** (Constant-equivalence). For all  $x \in \mathcal{F}$  there exists  $c \in C$  such that  $x \sim c$ .

**Axiom 6** (Finite Continuity). For any  $T$ , the restriction of  $\succ$  to  $X_T$  satisfies continuity, i.e., for any  $x \in X_T$  the sets  $\{y \in X_T : y \succ x\}$  and  $\{y \in X_T : y \prec x\}$  are open.

**Axiom 7** (Tail-continuity). For any  $c \in C$  and any  $x \in \mathcal{F}$  if  $x \succ c$ , then there exists  $\tau$  such that for all  $T \geq \tau$ ,  $x_T c \succ c$ ; if  $x \prec c$ , then there exists  $\tau$  such that for all  $T \geq \tau$ ,  $x_T c \prec c$ .

**Theorem 1.** *The preference  $\succsim$  satisfies Axioms 1–7 if and only if it is represented by (1.3) for some nonconstant and continuous  $u, v : X \rightarrow \mathbb{R}$  and  $0 < \delta < 1$ .*

Note that the representation obtained in Theorem 1 is a generalization of the quasi-hyperbolic model. The main two features of this representation are the intertemporal

separability of consumption and the standard stationary behavior that follows period 1 (captured by the quasi-stationarity axiom). The restriction that specifies representation (1.3) to the quasi-hyperbolic class imposes a strong relationship between the utility functions  $u$  and  $v$ . Not only do they have to represent the same ordering over the consumption space  $C$ , but also they must preserve the same cardinal ranking, i.e.  $u$  and  $v$  relate to each other through a positive affine transformation  $u = \beta v$  (the additive constant can be omitted without loss of generality). In order to capture this restriction behaviorally we express it in terms of the willingness to make tradeoffs between time periods.

We now present three different ways of restricting (1.3) to the quasi-hyperbolic model. It is important to observe that an axiom that requires the preference relation  $\succsim$  to exhibit preference pattern (1.1) is necessary, but not sufficient to pin down the  $\beta\delta$  model: present bias may arise as an immediate consequence of different preference intensity—as captured by differences in  $u$  and  $v$ . Therefore, in the context of representation (1.3), present bias could be explained without relying on the  $\beta\delta$  structure. The additional axioms that we propose, shed light on what it exactly means, in terms of consumption behavior, to have different short term discount factors and a common utility index.

### 1.3.1 Compensation Axiom

First, we present an axiom that ensures  $\delta$  is larger than half. We impose this requirement in order to be able to construct a “future compensation scheme” that exactly offsets the lengthening of the first time period caused by  $\beta$ . If  $\delta$  is less than half, then there will be values of  $\beta$  which we cannot compensate for exactly.<sup>4</sup>

---

<sup>4</sup>Since in most calibrations  $\delta$  is close to one for any reasonable length of the time period, we view this step as innocuous.

**Axiom 8** ( $\delta \geq 0.5$ ). If  $(c, a, a, \dots) \succ (c, b, b, \dots)$  for some  $a, b, c \in X$ , then

$$(c, b, a, a, \dots) \succsim (c, a, b, b, \dots).$$

In the context of representation (1.3) the long-run patience ( $\delta$ ) can be easily measured. Fix two elements  $a, b \in C$  such that  $a$  is preferred to  $b$ . Axiom 8 uncovers the strength of patience by getting information about the following tradeoff. Consider first a consumption stream that pays  $a$  tomorrow and  $b$  forever after. Consider now a second consumption stream in which the order of the alternatives is reversed. An agent that decides to postpone higher utility (by choosing  $b$  first) reveals a certain degree of patience. Under representation (1.3) the patient choice reveals a value of  $\delta \geq .5$ .

**Theorem 2.** *Suppose  $\succsim$  is as in Theorem 1. It satisfies Axiom 8 if and only  $\delta \geq 0.5$ .*

As discussed in the Introduction, our main axiom relies on the idea of increasing the distance between future payoffs to compensate for the lengthening of the time horizon caused by  $\beta$ . For example, if  $\beta = \delta$ , then the tradeoff between periods 0 and 1 is the same as the tradeoff between periods 1 and 3. Similarly, if  $\beta = \delta^t$ , then the tradeoff between periods 0 and 1 is the same as the tradeoff between periods 1 and  $t+2$ . Because we are working in discrete time, there exist values of  $\beta$  such that  $\delta^{t+1} < \beta < \delta^t$  for some  $t$ , so that the exact compensation of this form is not possible. However, due to time separability, other forms of compensation will be exact. Lemma 1 in the Appendix shows that as long as  $\delta \geq 0.5$ , any value of  $\beta$  can be represented by a sum of the powers of  $\delta$ .<sup>5</sup> The set  $M$  captures these powers; formally, let  $M$  denote a subset of  $\{2, 3, \dots\} \subseteq \mathcal{T}$ . We will refer to  $M$  as *compensation*. Our main axiom guarantees that the set  $M$  is independent of the consumption levels used to elicit the tradeoffs.

---

<sup>5</sup>A similar technique was used in repeated games, see, e.g., Sorin (1986) and Fudenberg and Maskin (1991). We thank Drew Fudenberg for these references.

**Axiom 9** (Compensation). There exists a compensation  $M$  such that for all  $a, b, c, d, e$

$$\begin{pmatrix} a & \text{if } t = 0 \\ b & \text{if } t = 1 \\ e & \text{otherwise} \end{pmatrix} \succ \begin{pmatrix} c & \text{if } t = 0 \\ d & \text{if } t = 1 \\ e & \text{otherwise} \end{pmatrix}$$

if and only if

$$\begin{pmatrix} a & \text{if } t = 1 \\ b & \text{if } t \in M \\ e & \text{otherwise} \end{pmatrix} \succ \begin{pmatrix} c & \text{if } t = 1 \\ d & \text{if } t \in M \\ e & \text{otherwise} \end{pmatrix}.$$

The main result of our paper is the following theorem.

**Theorem 3.** *A preference  $\succsim$  satisfies Axioms 1–9 if and only if has a quasi-hyperbolic discounting representation with  $\delta \geq 0.5$ . In this case,  $\beta = \sum_{t \in M} \delta^{t-2}$ .*

### 1.3.2 Alternate Approaches

The compensation axiom ensures that  $v$  is cardinally equivalent to  $u$ . From the formal logic viewpoint, however, the compensation axiom involves an existential quantifier. This section complements our analysis by considering two alternate ways of ensuring the cardinal equivalence: a form of the tradeoff consistency axiom and a form of the independence axiom.

Both axioms need to be complemented with an axiom that guarantees that  $\beta < 1$ . The following axiom yields just that.

**Axiom 10** (Present Bias). For any  $a, b, c, d, e \in C$ ,  $a \succ c$

$$(e, a, b, e, \dots) \sim (e, c, d, e, \dots) \implies (a, b, e, \dots) \succsim (c, d, e, \dots).$$

This axiom says that if two distant consumption streams are indifferent, one “impatient” (involving a bigger prize at  $t = 1$ , followed by a smaller at  $t = 2$ ) and one

“patient” (involving a smaller prize at  $t = 1$ , followed by a bigger at  $t = 2$ ), then pushing both of them forward will skew the preference toward the “impatient” choice.

For both approaches, fix a consumption level  $e \in C$  (for example in the context of monetary prizes,  $e$  could be zero dollars). For any pair of consumption levels  $a, b \in C$  let  $(a, b)$  denote the consumption stream  $(a, b, b, b, \dots)$ .

### Tradeoff Consistency Axiom

**Axiom 11** (Tradeoff Consistency). For any  $a, b, c, d, e_1, e_2 \in C$ ,

If  $(b, e_2) \succsim (a, e_1)$ ,  $(c, e_1) \succsim (d, e_2)$ , and  $(e_3, a) \sim (e_4, b)$ , then  $(e_3, c) \succsim (e_4, d)$ .

and

If  $(e_2, b) \succsim (e_1, a)$ ,  $(e_1, c) \succsim (e_2, d)$ , and  $(a, e_3) \sim (b, e_4)$ , then  $(c, e_3) \succsim (d, e_4)$ .

The intuition behind the first requirement of axiom is as follows (the second requirement is analogous and ensures that the time periods are being treated symmetrically). The first premise is that the “utility difference” between  $b$  and  $a$  offsets the utility difference between  $e_1$  and  $e_2$ . The second premise is that the utility difference between  $e_1$  and  $e_2$  offsets the utility difference between  $d$  and  $c$ . These two taken together imply that the utility difference between  $b$  and  $a$  is bigger than the utility difference between  $d$  and  $c$ . Thus, if the utility difference between  $e_3$  and  $e_4$  exactly offsets the utility difference between  $b$  and  $a$ , it must be big enough to offset the utility difference between  $d$  and  $c$ .

**Theorem 4.** *The preference  $\succsim$  satisfies Axioms 1–7 and 11 if and only if there exists a nonconstant and continuous function  $u : C \rightarrow \mathbb{R}$  and parameters  $\beta > 0$  and  $\delta \in (0, 1)$  such that  $\succsim$  is represented by the mapping*

$$x \mapsto u(x_0) + \beta \sum_{t=1}^{\infty} \delta^t u(x_t).$$

Moreover, it satisfies Axiom 10 if and only if  $\beta \leq 1$ , i.e.,  $\succsim$  has the quasi-hyperbolic discounting representation.

### Independence Axiom

By continuity (Axioms 6 and 7) for any  $a, b \in C$  there exists a consumption level  $c$  that satisfies  $(c, c) \sim (a, b)$ . Let  $c(a, b)$  denote the set of such consumption levels. Note that we are not imposing any monotonicity assumptions on preferences (the set  $C$  could be multidimensional) and for this reason the set  $c(a, b)$  may not be a singleton. However, since all of its members are indifferent to each other, it is safe to assume in the expressions below that  $c(a, b)$  is an arbitrarily chosen element of that set.

**Axiom 12** (Independence). For any  $a, a', a'', b, b', b'' \in C$  if  $(a, b) \succsim (a', b')$ , then

$$(c(a, a''), c(b, b'')) \succsim (c(a', a''), c(b', b''))$$

and

$$(c(a'', a), c(b'', b)) \succsim (c(a'', a'), c(b'', b')).$$

The intuition behind the first requirement of the axiom is as follows (the second requirement is analogous and ensures that the time periods are being treated symmetrically): For any  $(a, b), (a'', b'')$  the stream given by  $(c(a, a''), c(b, b''))$  is a “subjective mixture” of bets  $(a, b)$  and  $(a'', b'')$ . The axiom requires that if one consumption stream is preferred to another, then mixing each stream with a third stream preserves the preference.<sup>6</sup>

The next axiom, is a version of Savage’s P3. It ensures that preferences in each time period are ordinally the same.

---

<sup>6</sup>We thank Simon Grant for suggesting this type of axiom. A similar approach along the lines of Nakamura (1990) is considered in the Appendix.



**Axiom 13.** (Monotonicity) For any  $a, b, e \in C$ , then

$$b \succsim a \iff (b, e) \succsim (a, e) \text{ and } (e, b) \succsim (e, a)$$

**Theorem 5.** *The preference  $\succsim$  satisfies Axioms 1–7 and 12–13 if and only if there exists a nonconstant and continuous function  $u : C \rightarrow \mathbb{R}$  and parameters  $\beta > 0$  and  $\delta \in (0, 1)$  such that  $\succsim$  is represented by the mapping*

$$x \mapsto u(x_0) + \beta \sum_{t=1}^{\infty} \delta^t u(x_t).$$

*Moreover, it satisfies Axiom 10 if and only if  $\beta \leq 1$ , i.e.,  $\succsim$  has the quasi-hyperbolic discounting representation.*

## 1.4 Experimental Design

The axiom system presented in the previous section is suggestive of a new experimental design. The proposed experiment provides a direct test of stationarity; moreover, under the assumption that the preference belongs to the quasi-hyperbolic class, the experiment yields two sided bounds on the discount factors  $\beta$  and  $\delta$ .

The size of the bounds can be controlled by the appropriate choice of the compensation  $M$ . In the design proposed here we use the simplest compensation composed of just two consecutive elements, but more accurate measurements are possible.

For any choice of  $M$  the experiment does not rely on any assumptions about the curvature of the utility function  $u$ . In fact, whether the prizes are monetary or not is immaterial; the only assumption that the researcher has to make is that there exist two prizes  $a$  and  $b$ , where  $b$  is more preferred than  $a$  (but it doesn't matter “by how much”). As a consequence, the experimental design can be used to study how the nature of the prize (e.g., money, consumption good, addictive good) affects impatience, a feature not shared by experiments based on varying monetary payoffs.

The questionnaire involves a series of questions, each with the same structure. Each question asks for a comparison of two consumption plans: one “impatient” (with an early good payoff followed by two worse payoffs), the other “patient” (with an early worse payoff, followed by two good payoffs). No payoffs are given in any other periods. For example

$$\begin{pmatrix} b & \text{if } t = 1 \\ a & \text{if } t = 7 \\ a & \text{if } t = 8 \end{pmatrix} \quad \text{versus} \quad \begin{pmatrix} a & \text{if } t = 1 \\ b & \text{if } t = 7 \\ b & \text{if } t = 8 \end{pmatrix}$$

Under the assumption that the subject chooses according to the  $\beta$ - $\delta$  model, an answer to this question yields a bound on  $\delta$ , as the choice of the patient consumption plan implies

$$u(a) + (\delta^6 + \delta^7)u(b) \geq u(b) + (\delta^6 + \delta^7)u(a)$$

if and only if  $\delta^6 + \delta^7 \geq 1$  if and only if  $\delta \geq .89$ . Similarly, the comparison of

$$\begin{pmatrix} b & \text{if } t = 0 \\ a & \text{if } t = 1 \\ a & \text{if } t = 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a & \text{if } t = 0 \\ b & \text{if } t = 1 \\ b & \text{if } t = 2 \end{pmatrix}$$

yields a bound on  $\beta$ . For example, if the first alternative was chosen, then

$$u(a) + \beta(\delta + \delta^2)u(b) \leq u(b) + \beta(\delta + \delta^2)u(a)$$

if and only if  $\beta \leq \frac{1}{\delta + \delta^2}$ ; using the bound on  $\delta$  obtained above we have  $\beta \leq .59$ .

Stringing together a series of comparisons in a similar fashion with varying initial and secondary delays yields two sided bounds on  $\beta$  and  $\delta$ . Generally, each comparison

is indexed by two parameters: the initial delay  $T$  and the secondary delay  $S$ .

$$\begin{pmatrix} b & \text{if } t = T \\ a & \text{if } t = T + S \\ a & \text{if } t = T + S + 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a & \text{if } t = T \\ b & \text{if } t = T + S \\ b & \text{if } t = T + S + 1 \end{pmatrix}$$

For example, the first comparison had  $T = 1, S = 6$  and the second had  $T = 0, S = 1$ .

Varying  $T$  for a fixed  $S$  provides a direct test of stationarity and of quasi-stationarity. The comparison can also be helpful in detecting the correct period length—the minimal period length for which quasi-stationarity holds. Varying  $S$  together with  $T$  yields bounds on the discount factors, as exemplified above. This experimental design can be also used for parameter measurement under functional form assumptions other than  $\beta - \delta$ , as well as a test of behavioral properties that do not rely on any functional form assumptions; for example, varying  $T$  provides a direct test of stationarity.

### 1.4.1 Alternate Design

An alternate way of eliciting  $\beta$  that is suggested by Axiom 9. Suppose that  $\delta$  is known and we are interested in testing the hypothesis that  $\beta$  lies in some particular interval. For concreteness, suppose that  $\delta = .9$  and that we want to check whether  $\beta \in (.72, .81)$ . We use Axiom 9 to test this hypothesis. First, find prizes  $a \succ c, d \succ b$  such that

$$\begin{pmatrix} a & \text{if } t = 0 \\ b & \text{if } t = 1 \end{pmatrix} \sim \begin{pmatrix} c & \text{if } t = 0 \\ d & \text{if } t = 1 \end{pmatrix}$$

it can be shown that  $\beta < \delta^2$  if and only if:

$$\begin{pmatrix} a & \text{if } t = 1 \\ b & \text{if } t = 4 \end{pmatrix} \prec \begin{pmatrix} c & \text{if } t = 1 \\ d & \text{if } t = 4 \end{pmatrix}$$

and  $\beta > \delta^3$  if and only if

$$\begin{pmatrix} a & \text{if } t = 1 \\ b & \text{if } t = 5 \end{pmatrix} \succ \begin{pmatrix} c & \text{if } t = 1 \\ d & \text{if } t = 5 \end{pmatrix}.$$

Violations of any of the two previous relations will suggest that  $\beta \notin (\delta^3, \delta^2)$ .

## 1.5 Related Literature

A large part of the theoretical literature on time preferences uses the the choice domain of *dated rewards*, where preferences are defined on  $C \times \mathcal{T}$ , i.e., only one payoff is made. On this domain Fishburn and Rubinstein (1982) axiomatized exponential discounting. By assuming that  $\mathcal{T} = \mathbb{R}_+$ , i.e., that time is continuous, Loewenstein and Prelec (1992) axiomatized a generalized model of *hyperbolic discounting*, where preferences are represented by  $V(x, t) = (1 + \alpha t)^{-\frac{\beta}{\alpha}} u(x)$ . Recently, Attema *et al.* (2010) generalized this method and obtained an axiomatization of quasi-hyperbolic discounting, among other models.

The above results share a common problem: the domain of dated rewards is not rich enough to enable the measurement of the levels of discount factors. Even in the exponential discounting model the value of  $\delta$  can be chosen arbitrarily, as long as it belongs to the interval  $(0, 1)$ , see, e.g., Theorem 2 of Fishburn and Rubinstein (1982); see also the recent results of Noor (2011). The richer domain of consumption streams that we employ in this paper allows us to elicit more complex tradeoffs between time periods and to pin down the value of all discount factors.

The continuous time approach can be problematic for yet another reason. It relies on extracting a sequence of time periods of equal subjective length, a so called standard

sequence.<sup>7</sup> Since the time intervals in a standard sequence are of equal subjective length, their objective duration is unequal and has to be uncovered by eliciting indifference. In contrast, our method uses time intervals of objectively equal length and does not rely on such elicitation.

Finally, an axiomatization of quasi-hyperbolic discounting using a different approach was obtained by Hayashi (2003). He studied preferences over an extended domain that includes lotteries over consumption streams. He used the lottery mixtures to calibrate the value of  $\beta$ . His axiomatization and measurement rely heavily on the assumption of expected utility, which is rejected by the bulk of experimental evidence. Moreover, in his model the same utility function  $u$  measures both risk aversion and the intertemporal elasticity of substitution; however these two features of preferences are conceptually unrelated (see, e.g., Kreps and Porteus, 1978; Epstein and Zin, 1989b) and are shown to be different in empirical calibrations. Another limitation of his paper is that his axioms are not suggestive of a measurement method of the relation between the short-run and long-discount factor.

---

<sup>7</sup>The standard sequence method was originally applied to eliciting subjective beliefs by Ramsey (1926) and later by Luce and Tukey (1964). Interestingly, the similarity between beliefs and discounting was already anticipated by Ramsey: “the degree of belief is like a time interval; it has no precise meaning unless we specify how it is to be measured.”

# Chapter 2

## A Robust Test for Weak Instruments <sup>1</sup>

### 2.1 Introduction

This paper proposes a simple test for weak instruments that is robust to heteroskedasticity, serial correlation, and clustering. Staiger and Stock (1997) and Stock and Yogo (2005) developed widely used tests for weak instruments under the assumption of conditionally homoskedastic serially uncorrelated model errors. However, applications with heteroskedasticity, time series autocorrelation, and clustered panel data are common. Our proposed test provides empirical researchers with a new tool to assess instrument strength for those applications.

The practical relevance of heteroskedasticity in linear instrumental variable (IV) regression has been highlighted before by Antoine and Lavergne (2012), Chao and Newey (2012) and Hausman *et al.* (2012). We show, more generally, that departures from the conditionally homoskedastic serially uncorrelated framework affect the weak instrument asymptotic distribution of both the Two-Stage Least Squares (TSLS) and the Limited Information Maximum Likelihood (LIML) estimators. Consequently, heteroskedastic-

---

<sup>1</sup>Co-authored with Carolin Pflueger

ity, autocorrelation, and/or clustering can further bias estimators and distort test sizes when instruments are potentially weak. At the same time, the first stage may falsely indicate that instruments are strong.

Under strong instruments, both TSLS and LIML are asymptotically unbiased, while such is generally not the case when instruments are weak. We follow the standard Nagar (1959) methodology to derive a tractable proxy for the asymptotic estimator bias that is defined for both TSLS and LIML. Our procedure tests the null hypothesis that the Nagar bias is large relative to a “worst-case” benchmark. Our benchmark coincides with the Ordinary Least Squares (OLS) bias benchmark when the model errors are conditionally homoskedastic and serially uncorrelated, but differs otherwise.

Our proposed test statistic, which we call the effective F statistic, is a scaled version of the non-robust first stage F statistic. The null hypothesis for weak instruments is rejected for large values of the effective F. The critical values depend on an estimate of the covariance matrix of the OLS reduced form regression coefficients, and on the covariance matrix of the reduced form errors, which can be estimated using standard procedures.

We consider two different testing procedures, generalized and simplified; both are asymptotically valid. Critical values for both procedures can be calculated either by Monte-Carlo methods, or by a curve-fitting methodology by Patnaik (1949). The generalized testing procedure applies to both TSLS and LIML, and has increased power, but is computationally more demanding. In contrast, the simplified procedure applies only to TSLS. The simplified procedure is conservative, because it protects against the worst type of heteroskedasticity, serial correlation, and/or clustering in the second stage.

Empirical researchers frequently report the robust F statistic as a simple way of adjusting the Staiger and Stock (1997) and Stock and Yogo (2005) pre-tests for het-

eroskedasticity, serial correlation, and clustering, and compare them to the homoskedastic critical values. To the best of our knowledge, there is no theoretical or analytical support for this practice, as cautioned in Baum *et al.* (2007). Our proposed procedures adjust the critical values. While our proposed test statistic corresponds to the robust F statistic in the just identified case, it differs in the over-identified case.

Our baseline implementation tests the null hypothesis that the Nagar bias exceeds 10% of a “worst-case” bias with a size of 5%. The simplified procedure for TSLS has critical values between 11 and 23.1 that depend only on the covariance matrix of the first stage reduced form coefficients. Thus a simple, asymptotically valid rule of thumb is available for TSLS that rejects when the effective F is greater than 23.1.

We apply weak instrument pre-tests to a well-known empirical example, the IV estimation of the Elasticity of Intertemporal Substitution (EIS) (Yogo, 2004; Campbell, 2003). Our empirical results are consistent with Yogo (2004)’s finding that the EIS is small and close to zero. However, for several countries in our sample, conditionally homoskedastic serially uncorrelated pre-tests indicate strong instruments, while our proposed test cannot reject the null hypothesis of weak instruments.

There is a large literature on inference when IVs are weak; see Stock *et al.* (2002) and Andrews and Stock (2006) for overviews. Our paper is closest to Staiger and Stock (1997) and Stock and Yogo (2005). Zhan (2010) provides another interesting approach, which, unlike ours, proposes to test the null hypothesis of strong instruments. Also, Bun and de Haan (2010) point out the invalidity of pre-tests based on the first stage F statistic in two particular examples of non-homoskedastic and serially correlated errors, but do not provide a valid pre-test.

Robust methods for inference about the coefficients of a single endogenous regressor when IVs are weak and errors are heteroskedastic and/or serially correlated are also available (Andrews and Stock, 2006; Kleibergen, 2007). A pre-test for weak instru-



ments followed by standard inference procedures can be less computationally demanding, and the use of this two-stage decision rule is widespread because of its simplicity. We therefore view this paper as complementary to robust inference methods.

It is well-known that pre-tests can induce uniformity problems (Leeb and Poetscher, 2005; Guggenberger, 2010a,b). However, Stock and Yogo (2005) have shown that in the conditionally homoskedastic and serially uncorrelated case the first stage F statistic can be used to control the Wald test size distortion. In this case, uniformity problems are therefore not a first order concern: a two-stage test that uses the Wald test after observing a large effective F statistic has asymptotic size well below unity under weak instrument asymptotics.

The rest of the paper is organized as follows. Section 2 introduces the model, and presents the generalized and simplified testing procedures. Section 3 derives asymptotic distributions, and shows that conditional heteroskedasticity and serial correlation can effectively weaken instruments in an illustrative example. Section 4 derives the expressions for the TSLS and LIML Nagar biases, and describes the test statistic and critical values. Section 5 discusses the implementation of the critical values by Monte Carlo simulation and Patnaik (1949)'s methodology. Section 6 applies the pre-testing procedure to the IV estimation of the EIS. Section 7 concludes. All proofs are collected in the Appendix to this Chapter..

## 2.2 Model and Summary of Testing Procedure

### 2.2.1 Model and Assumptions

We consider a linear IV model in reduced form with one endogenous regressor and  $K$  instruments

$$\mathbf{y} = \mathbf{Z}\mathbf{\Pi}\beta + \mathbf{v}_1 \quad (2.1)$$

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Pi} + \mathbf{v}_2 \quad (2.2)$$

The structural parameter of interest is  $\beta \in \mathbb{R}$ , while  $\mathbf{\Pi} \in \mathbb{R}^K$  denotes the unknown first stage parameter vector. The sample size is  $S$  and the econometrician observes the data set  $\{y_s, Y_s, \mathbf{Z}_s\}_{s=1}^S$ . We denote observations of the outcome variable, the endogenous regressor, and the vector of instruments by  $y_s, Y_s$  and  $\mathbf{Z}_s$ , respectively. The unobserved reduced form errors have realizations  $v_{js}, j \in \{1, 2\}$ . We stack the realized variables in matrices  $\mathbf{y} \in \mathbb{R}^S$ ,  $\mathbf{Z} \in \mathbb{R}^{S \times K}$ , and  $\mathbf{v}_j \in \mathbb{R}^S, j \in \{1, 2\}$ .

Our analysis extends straightforwardly to a model with additional exogenous regressors. In the presence of additional exogenous regressors, TSLS and LIML estimators are unchanged if we replace all variables by their projection errors onto those exogenous regressors. TSLS and LIML are also invariant to normalizing the instruments to be orthonormal. We can therefore assume without loss of generality that there are no additional exogenous regressors, and that  $\mathbf{Z}'\mathbf{Z}/S = \mathbb{I}_K$ . When implementing the pre-test, an applied researcher needs to normalize the data.

We model weak instruments by assuming that the IV first stage relation is local to zero, following the modeling strategy in Staiger and Stock (1997).

**Assumption  $\mathbf{L}_{\mathbf{\Pi}}$ . (Local to Zero)**  $\mathbf{\Pi} = \mathbf{\Pi}_S = \mathbf{C}/\sqrt{S}$ , where  $\mathbf{C}$  is a fixed vector  $\mathbf{C} \in \mathbb{R}^K$ .

Additional high-level assumptions allow us to derive asymptotic distributions for

IV estimators and F statistics. TLS and LIML estimators and first stage F statistics depend on the statistics  $\mathbf{Z}'\mathbf{v}_j/\sqrt{S}$ , and estimates of the covariance matrices  $\mathbf{W}$  and  $\mathbf{\Omega}$  as defined below.

**Assumption HL. (High Level)** The following limits hold as  $S \rightarrow \infty$ .

1.  $\begin{pmatrix} \mathbf{Z}'\mathbf{v}_1/\sqrt{S} \\ \mathbf{Z}'\mathbf{v}_2/\sqrt{S} \end{pmatrix} \xrightarrow{d} \mathcal{N}_{2K}(\mathbf{0}, \mathbf{W})$  for some positive definite  $\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{W}_{12} \\ \mathbf{W}_{12}' & \mathbf{W}_2 \end{pmatrix}$
2.  $[\mathbf{v}_1, \mathbf{v}_2]'[\mathbf{v}_1, \mathbf{v}_2]/S \xrightarrow{p} \mathbf{\Omega}$  for some positive definite  $\mathbf{\Omega} \equiv \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix}$
3. There exists a sequence of positive definite estimates  $\{\widehat{\mathbf{W}}(S)\}$ , measurable with respect to  $\{y_s, Y_s, \mathbf{Z}_s\}_{s=1}^S$ , such that  $\widehat{\mathbf{W}}(S) \xrightarrow{p} \mathbf{W}$  as  $S \rightarrow \infty$

Assumption HL is satisfied under various primitive conditions on the joint distribution of  $(\mathbf{Z}, \mathbf{v}_1, \mathbf{v}_2)$ ; see Supplementary Materials C.2 for examples. Assumption HL.1 is satisfied as long as a Central Limit Theorem holds for  $\mathbf{Z}'\mathbf{v}_j/\sqrt{S}$ . Assumption HL.2 holds under a Weak Law of Large Numbers for  $[\mathbf{v}_1, \mathbf{v}_2]'[\mathbf{v}_1, \mathbf{v}_2]/S$ . Assumption HL.3 assumes that we can consistently estimate the covariance matrix  $\mathbf{W}$  from the observable variables.

Assumption HL allows for a general form of  $\mathbf{W}$ , similarly to the models in Müller (2011) and Mikusheva (2010). This is our key generalization from the model proposed in Staiger and Stock (1997), who require  $\mathbf{W}$  to have the form  $\mathbf{\Omega} \otimes I_K$ . The Kronecker form arises naturally only in the context of a conditionally homoskedastic serially uncorrelated model. Our generalization is therefore relevant for practitioners working with heteroskedastic, time series, or panel data, and it is consequential for econometric practice.

## 2.2.2 Implementing the Testing Procedure

### 2.2.1 Generalized Test

The generalized testing procedure can be implemented in four simple steps. When rejecting the null, the empirical researcher can conclude that the estimator Nagar bias is small relative to the benchmark. Under the null hypothesis, the Nagar bias of TSLS or LIML is greater than a fraction  $\tau$  of the benchmark. Critical values for the effective F statistic depend on the desired threshold  $\tau$ , the desired level of significance  $\alpha$ , and estimates for the matrices  $\widehat{\mathbf{W}}$ ,  $\widehat{\mathbf{\Omega}}$ . Critical values also vary between TSLS and LIML. In our numerical results, we focus on  $\tau = 10\%$  and  $\alpha = 5\%$ .

1. If there are additional exogenous regressors, replace all variables by their projection residuals onto those exogenous regressors. Normalize instruments to be orthonormal.
2. Obtain  $\widehat{\mathbf{W}}$  as the estimate for the asymptotic covariance matrix of the reduced form OLS coefficients. Standard statistical packages estimate this matrix (divided by the sample size  $S$ ) under different distributional assumptions. For cross-sectionally heteroskedastic applications, use a heteroskedasticity robust estimate; for time series applications, use a heteroskedasticity and autocorrelation consistent (HAC) estimate; and for panel data applications, use a “clustered” estimate.
3. Compute the test statistic, the *Effective F Statistic*

$$\widehat{F}_{eff} \equiv \frac{1}{S} \frac{\mathbf{Y}'\mathbf{Z}\mathbf{Z}'\mathbf{Y}}{tr(\widehat{\mathbf{W}}_2)} \quad (2.3)$$

where  $tr(\cdot)$  denotes the trace operator.

4. Estimate the *effective degrees of freedom*

$$\hat{K}_{eff} \equiv \frac{\left[ \text{tr} \left( \widehat{\mathbf{W}}_2 \right) \right]^2 (1 + 2x)}{\text{tr} \left( \widehat{\mathbf{W}}_2' \widehat{\mathbf{W}}_2 \right) + 2x \text{tr} \left( \widehat{\mathbf{W}}_2 \right) \max \text{eval}(\widehat{\mathbf{W}}_2)} \quad (2.4)$$

$$\text{where } x = B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau \text{ for } e \in \{\text{TSLS, LIML}\} \quad (2.5)$$

Here,  $\max \text{eval}(\widehat{\mathbf{W}}_2)$  denotes the maximum eigenvalue of the lower diagonal  $K \times K$  block of the matrix  $\widehat{\mathbf{W}}$ . The function  $B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})$  is closely related to the supremum of the Nagar bias relative to the benchmark; see Theorem 1.2. The numerical implementation of  $B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})$  is discussed in Remark 5, Theorem 1. A fast numerical MATLAB routine is available for the function  $B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})$ .

The generalized test rejects the null hypothesis of weak instruments when  $\hat{F}_{eff}$  exceeds a critical value that can be obtained by either of the following procedures:

- a) Monte Carlo methods, as described in Section 5;
- b) Patnaik (1949)'s curve-fitting methodology; Patnaik critical values obtain as the upper  $\alpha$  quantile of  $\chi^2_{\hat{K}_{eff}}(x\hat{K}_{eff})/\hat{K}_{eff}$ , where  $\chi^2_{\hat{K}_{eff}}(x\hat{K}_{eff})$  denotes a non-central  $\chi^2$  distribution with  $\hat{K}_{eff}$  degrees of freedom and noncentrality parameter  $x\hat{K}_{eff}$ . Table 2.1 tabulates 5% Patnaik critical values.

### 2.2.2 Simplified Test

A simplified conservative version of the test is available for TSLS. The simplified procedure follows the same steps, but sets  $x = 1/\tau$  in Step 4. For a given effective degrees of freedom  $\hat{K}_{eff}$ , the simplified 5% critical value can be conveniently read off Table 2.1. For instance, the critical value for a threshold  $\tau = 10\%$  can be found in the column with  $x = 10$ . The simplified test does not require numerical evaluation of  $B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})$ ,

**Table 2.1:** *Critical Values***Upper 5% Quantile of  $\chi^2_{K_{eff}}(xK_{eff})/K_{eff}$** 

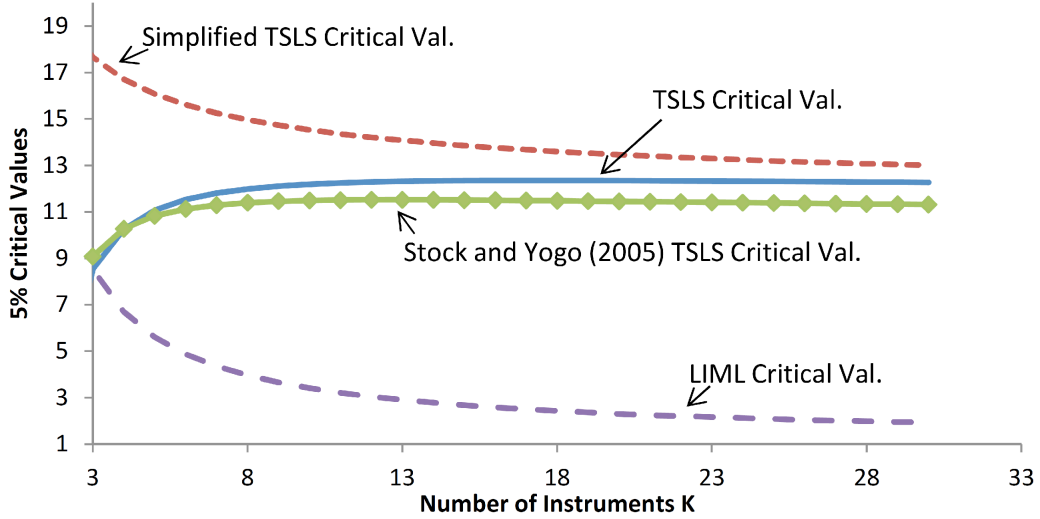
$K_{eff}$	$x = 3.33$	$x = 5$	$x = 10$	$x = 20$
1	12.05	15.06	23.11	37.42
2	9.57	12.17	19.29	32.32
3	8.53	10.95	17.67	30.13
4	7.92	10.23	16.72	28.85
5	7.51	9.75	16.08	27.98
6	7.21	9.40	15.62	27.35
7	6.98	9.14	15.26	26.86
8	6.80	8.92	14.97	26.47
9	6.65	8.74	14.73	26.15
10	6.52	8.59	14.53	25.87
11	6.41	8.47	14.36	25.64
12	6.32	8.36	14.21	25.44
13	6.24	8.26	14.08	25.26
14	6.16	8.17	13.96	25.10
15	6.10	8.10	13.86	24.96
16	6.04	8.03	13.77	24.83
17	5.99	7.96	13.68	24.71
18	5.94	7.91	13.60	24.60
19	5.89	7.85	13.53	24.50
20	5.85	7.80	13.46	24.41
21	5.81	7.76	13.40	24.33
22	5.78	7.72	13.35	24.25
23	5.74	7.68	13.29	24.18
24	5.71	7.64	13.24	24.11
25	5.68	7.61	13.20	24.05
26	5.66	7.57	13.15	23.98
27	5.63	7.54	13.11	23.93
28	5.61	7.51	13.07	23.87
29	5.58	7.49	13.04	23.82
30	5.56	7.46	13.00	23.77

NOTE: Critical values computed by Patnaik (1949) method. For generalized and simplified testing procedures, estimate  $K_{eff}$  as in (2.4). For a Nagar bias threshold  $\tau$  (e.g.  $\tau = 10\%$ ) set  $x = 1/\tau$  for the simplified procedure. For the generalized procedure, set  $x = B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau$ ; see Step 4 in Section 2.2.1.

for it uses the bound  $B_{TSLS}(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}}) \leq 1$ , proved in Theorem 1.3. The matrix  $\widehat{\mathbf{W}}$  enters only through the lower  $K \times K$  block  $\widehat{\mathbf{W}}_2$ .

### 2.2.3 Comparison with Stock and Yogo (05) Critical Values

We compare the generalized and simplified TSLS critical values to those proposed in the paper of Stock and Yogo (2005) for the case when the data is conditionally homoskedastic and serially uncorrelated. For this comparison, we assume  $\mathbf{W} = \mathbf{\Omega} \otimes \mathbb{I}_K$  and  $\mathbf{W}$  and  $\mathbf{\Omega}$  known, as in Stock and Yogo (2005). It then follows from (2.3) and (2.4) that the effective and non-robust F statistics are equal, and that the effective number of degrees of freedom  $K_{eff}$  equals the number of instruments  $K$ .



**Figure 2.1:** *TSLS and simplified 5% critical values*

Figure 1 shows the 5% TSLS critical value for testing the null hypothesis that the asymptotic estimator bias exceeds 10% of the benchmark, the 5% critical value for the corresponding simplified test, and the Stock and Yogo (2005) 5% critical value for testing the null hypothesis that the TSLS bias exceeds 10% of the OLS bias. The

Stock and Yogo (2005) critical value is defined when the degree of over identification is at least two and we therefore show critical values for  $3 \leq K \leq 30$ . The TSLS critical value increases from 8.53 for  $K = 3$  to 12.27 for  $K = 30$ . By comparison, the Stock and Yogo (2005) critical value increases from 9.08 for  $K = 3$  to 11.32 for  $K = 30$ . The simplified TSLS critical value is strictly larger than the TSLS critical value for all  $K$  shown, illustrating that the simplified test can be strictly less powerful than the generalized procedure. The difference between the simplified critical value and the TSLS and Stock and Yogo (2005) critical values decreases as  $K$  becomes large.

## 2.3 Asymptotic Distributions and an Example

### 2.3.1 Illustrative Example

A simple example illustrates that heteroskedasticity and serial correlation impact the entire asymptotic distribution of both TSLS and LIML estimators, and can weaken the performance of the estimators. In this example, the first stage F statistic rejects the null hypothesis of weak instruments too often, while the effective F statistic allows for testing for weak instruments with asymptotically correct size.

For the sake of exposition, assume  $\beta = 0$ . Also assume that the departure from the conditionally homoskedastic serially uncorrelated framework takes the particularly simple form

$$\mathbf{W} = a^2(\boldsymbol{\Omega} \otimes I_K) \tag{2.6}$$

$a$  is a scalar parameter and for  $a = 1$  the expression (2.6) reduces to the conditionally homoskedastic case.

**Remark 1.** We can generate example (2.6) with a purely conditionally heteroskedastic data-generating process. Let  $\{\mathbf{Z}_s, \tilde{v}_{1s}, \tilde{v}_{2s}\}$  identically and independently distributed



(i.i.d.). Let instruments independent with  $\mathbb{E}[Z_{ks}] = 0$ ,  $\mathbb{E}[Z_{ks}^2] = 1$ ,  $\mathbb{E}[Z_{ks}^3] = 0$ ,  $\mathbb{E}[Z_{ks}^4] = a^2$ . Let  $(\tilde{v}_{1s}, \tilde{v}_{2s}) \sim N_2((0, 0)', \mathbf{\Omega})$  independently of  $\mathbf{Z}_s$ . Let the reduced form errors  $v_{1s} = \tilde{v}_{1s}\Pi_{k=1}^K Z_{ks}$ ,  $v_{2s} = \tilde{v}_{2s}\Pi_{k=1}^K Z_{ks}$ . Then  $\mathbb{E}([v_{1s}, v_{2s}][v_{1s}, v_{2s}]') = \mathbf{\Omega}$  and  $\mathbb{E}([v_{1s}, v_{2s}][v_{1s}, v_{2s}]' \otimes \mathbf{Z}_s \mathbf{Z}_s') = a^2 \mathbf{\Omega} \otimes \mathbb{I}_K$ . HL.1, HL.2, and (2.6) follow from the Central Limit Theorem and the Weak Law of Large Numbers.

**Remark 2.** We can alternatively generate (2.6) with a simple serially correlated data-generating process. Assume that instruments and reduced form errors follow independent AR(1) processes  $Z_{ks+1} = \rho_Z Z_{ks} + \epsilon_{ks+1}$ ,  $k = 1, \dots, K$  and  $v_{js+1} = \rho_v v_{js} + \eta_{js+1}$ ,  $j = 1, 2$ . Let  $\epsilon_{ks}$  and  $\eta_{js}$  serially uncorrelated with mean zero,  $\mathbb{E}(\epsilon_s \epsilon_s') = (1 - \rho_Z^2) \times I_K$  and  $\mathbb{E}[\eta_{1s}, \eta_{2s}]'[\eta_{1s}, \eta_{2s}] = (1 - \rho_v^2) \times \mathbf{\Omega}$ . Then  $\mathbb{E}[v_{1s}, v_{2s}][v_{1s}, v_{2s}]' = \mathbf{\Omega}$  and  $\mathbb{E}(\mathbf{Z}_s \mathbf{Z}_s') = \mathbb{I}_K$ . HL.1, HL.2 follow from the Central Limit Theorem and the Weak Law of Large Numbers. Expression (2.6) holds with  $a = (1 + \rho_v \rho_Z)/(1 - \rho_Z \rho_v)$ . Serial correlation in both the instruments and the errors is required for  $a \neq 1$ . As a numerical example, moderate serial correlation of  $\rho_v = \rho_Z = 0.5$  gives rise to  $a = 1.67$ .

With Assumptions  $L_{\Pi}$  and HL the asymptotic distribution of the TSLS estimator

$$\hat{\beta}_{TSLS} \equiv [\mathbf{Y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}]^{-1}\mathbf{Y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v}_1 \quad (2.7)$$

$$= \frac{\omega_1}{\omega_2} \left[ \left( \frac{\mathbf{C}}{a\omega_2} + \frac{\mathbf{Z}'\mathbf{v}_2/\sqrt{S}}{a\omega_2} \right)' \left( \frac{\mathbf{C}}{a\omega_2} + \frac{\mathbf{Z}'\mathbf{v}_2/\sqrt{S}}{a\omega_2} \right) \right]^{-1} \quad (2.8)$$

$$\times \left( \frac{\mathbf{C}}{a\omega_2} + \frac{\mathbf{Z}'\mathbf{v}_2/\sqrt{S}}{a\omega_2} \right)' \frac{\mathbf{Z}'\mathbf{v}_1/\sqrt{S}}{a\omega_1} \quad (2.9)$$

$$\xrightarrow{d} \frac{\omega_1}{\omega_2} [\boldsymbol{\psi}_2' \boldsymbol{\psi}_2]^{-1} \boldsymbol{\psi}_2' \boldsymbol{\psi}_1 \quad (2.10)$$

where  $\begin{pmatrix} \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 \end{pmatrix} \sim \mathcal{N}_{2K} \left( \begin{pmatrix} \mathbf{0}_K \\ \mathbf{C}/(a\omega_2) \end{pmatrix}, \begin{pmatrix} 1 & \omega_{12}/(\omega_1\omega_2) \\ \omega_{12}/(\omega_1\omega_2) & 1 \end{pmatrix} \otimes \mathbb{I}_K \right).$

The asymptotic TSLS distribution depends only on the elements of the non-central Wishart matrix  $[\boldsymbol{\psi}_1, \boldsymbol{\psi}_2]'[\boldsymbol{\psi}_1, \boldsymbol{\psi}_2]$ . Hence, the vector of first stage coefficients  $\mathbf{C}$  and

the parameter  $a$  enter into the asymptotic distribution in (2.10) only through the noncentrality parameter  $\mathbf{C}'\mathbf{C}/a^2\omega_2^2$ , so  $\mathbf{C}'\mathbf{C}/a^2\omega_2^2$  summarizes instrument strength.

In this example, heteroskedasticity and serial correlation affect the biases and test size distortion of TSLS and LIML estimators in the same way as a weaker first stage relationship. The conditionally homoskedastic serially uncorrelated case obtains for  $a = 1$ , so the TSLS estimator is asymptotically distributed as if the errors were conditionally homoskedastic serially uncorrelated, and the first stage coefficients were reduced by a factor of  $a$ . We prove an analogous result for LIML in the Appendix to this Chapter..

Consider a null hypothesis for weak instruments of the form  $(\mathbf{C}'\mathbf{C}/\omega_2^2 a^2 K) < x$ . In the presence of conditional heteroskedasticity or serial correlation of the form (2.6), the first stage F statistic is asymptotically distributed as  $a^2 \chi_K^2 (\mathbf{C}'\mathbf{C}/\omega_2^2 a^2) / K$ . As  $a$  increases without bound, the noncentrality parameter goes to zero and instruments become arbitrarily weak, but the first stage F statistic diverges to infinity almost surely. On the other hand, the effective F statistic is asymptotically distributed as a  $\chi_K^2 (\mathbf{C}'\mathbf{C}/\omega_v^2 a^2) / K$ , so we can reject the null hypothesis of weak instruments with confidence level  $\alpha$  whenever  $\hat{F}_{eff}$  exceeds the upper  $\alpha$  quantile of  $\chi_K^2 (x \times K) / K$ .

### 2.3.2 Asymptotic Distributions

**Definition.** Denote the projection matrix onto  $\mathbf{Z}$  by  $\mathbf{P}_Z = \mathbf{Z}\mathbf{Z}'/S$  and the complementary matrix by  $\mathbf{M}_Z = \mathbb{I}_S - \mathbf{P}_Z$ .

1. The *Two-Stage Least Squares* (TSLS) estimator

$$\hat{\beta}_{TSLS} \equiv (\mathbf{Y}'\mathbf{P}_Z\mathbf{Y})^{-1}(\mathbf{Y}'\mathbf{P}_Z\mathbf{y}) \quad (2.11)$$

2. The *Limited Information Likelihood* (LIML) estimator

$$\hat{\beta}_{LIML} = (\mathbf{Y}'(\mathbb{I}_S - k_{LIML}\mathbf{M}_Z)\mathbf{Y})^{-1}(\mathbf{Y}'(\mathbb{I}_S - k_{LIML}\mathbf{M}_Z)\mathbf{y}) \quad (2.12)$$

where  $k_{LIML}$  is the smallest root of the determinantal equation

$$\left| [\mathbf{y}, \mathbf{Y}]'[\mathbf{y}, \mathbf{Y}] - k[\mathbf{y}, \mathbf{Y}]'\mathbf{M}_{\mathbf{Z}}[\mathbf{y}, \mathbf{Y}] \right| = 0 \quad (2.13)$$

3. The *non-robust first stage F statistic*

$$\widehat{F} \equiv \frac{\mathbf{Y}'\mathbf{P}_{\mathbf{Z}}\mathbf{Y}}{K \widehat{\omega}_2^2} \quad (2.14)$$

$$\text{where } \widehat{\omega}_2^2 \equiv \frac{(\mathbf{Y} - \mathbf{P}_{\mathbf{Z}}\mathbf{Y})'(\mathbf{Y} - \mathbf{P}_{\mathbf{Z}}\mathbf{Y})}{S - K - 1}$$

4. The *robust first stage F statistic*

$$\widehat{F}_r \equiv \frac{\mathbf{Y}'\mathbf{Z}\widehat{\mathbf{W}}_2^{-1}\mathbf{Z}'\mathbf{Y}}{K \times S} \quad (2.15)$$

where  $\widehat{\mathbf{W}}_2$  is the lower diagonal  $K \times K$  block of the matrix  $\widehat{\mathbf{W}}$ .

5. The *effective first stage F statistic*

$$\widehat{F}_{eff} \equiv \frac{\mathbf{Y}'\mathbf{P}_{\mathbf{Z}}\mathbf{Y}}{\text{tr}(\widehat{\mathbf{W}}_2)} \quad (2.16)$$

Lemma 1 derives asymptotic distributions for these statistics, generalizing Theorem 1 in Staiger and Stock (1997).

**Lemma 1.** Write  $\sigma_1^2 = \omega_1^2 - 2\beta\omega_{12} + \beta^2\omega_2^2$ ,  $\sigma_{12} = \omega_{12} - \beta\omega_2^2$ ,  $\sigma_2^2 = \omega_2^2$  and  $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ . Under Assumptions  $L_{\Pi}$  and  $HL$  the following limits hold jointly as  $S \rightarrow \infty$ .

1.  $\widehat{\beta}_{TSLS} - \beta \xrightarrow{d} \beta_{TSLS}^* = (\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2)^{-1} \boldsymbol{\gamma}_2'(\boldsymbol{\gamma}_1 - \beta\boldsymbol{\gamma}_2)$
2.  $\widehat{\beta}_{LIML} - \beta \xrightarrow{d} \beta_{LIML}^* = (\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2 - \kappa_{LIML}\omega_2^2)^{-1} (\boldsymbol{\gamma}_2'(\boldsymbol{\gamma}_1 - \beta\boldsymbol{\gamma}_2) - \kappa_{LIML}(\omega_{12} - \beta\omega_2^2))$

where  $\kappa_{LIML}$  is the smallest root of  $|[\boldsymbol{\gamma}_1 - \beta\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_2]'[\boldsymbol{\gamma}_1 - \beta\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_2] - \kappa\boldsymbol{\Sigma}| = 0$

3.  $\widehat{F} \xrightarrow{d} F^* \equiv \gamma_2' \gamma_2 / K \omega_2^2$
4.  $\widehat{F}_r \xrightarrow{d} F_r^* \equiv \gamma_2' \mathbf{W}_2^{-1} \gamma_2 / K$
5.  $\widehat{F}_{eff} \xrightarrow{d} F_{eff}^* \equiv \gamma_2' \gamma_2 / \text{tr}(\mathbf{W}_2)$

Where

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \sim \mathcal{N}_{2K} \left( \begin{pmatrix} \beta \mathbf{C} \\ \mathbf{C} \end{pmatrix}, \mathbf{W} \right) \quad (2.17)$$

*Proof.* See the Appendix to this Chapter.. □

The limiting distributions are functions of a multivariate normal vector whose distribution depends on the parameters  $(\beta, \mathbf{C})$ , and on the matrix  $\mathbf{W}$ . We treat the asymptotic distributions in Lemma 1 as a limiting experiment in the sense of Müller (2011), and use it to analyze inference problems regarding  $(\beta, \mathbf{C})$ .

## 2.4 Testing the Null Hypothesis of Weak Instruments

We base our null hypothesis of weak instruments on a bias criterion. We follow the standard methodology in Nagar (1959), and approximate the asymptotic TLSL and LIML distributions to obtain the Nagar bias. Under standard asymptotics, the Nagar bias for both estimators is zero everywhere in the parameter space, but under weak instrument asymptotics, the bias may be large in some regions of the parameter space. We consider instruments to be weak when the estimator Nagar bias is large relative to a benchmark, extending the OLS bias criterion in Stock and Yogo (2005).

### 4.1 Nagar Approximation

**Theorem 1. (*Nagar Approximation*)** Let  $\mathbf{W} \in \mathbb{R}^{2K \times 2K}$  positive definite. Write  $\mathbf{C} \in \mathbb{R}^K$  as  $\mathbf{C} = \|\mathbf{C}\| \mathbf{C}_0$  and let  $\mu^2 \equiv \|\mathbf{C}\|^2 / \text{tr}(\mathbf{W}_2)$ . Define  $\mathbf{S}_1 = \mathbf{W}_1 - 2\beta \mathbf{W}_{12} + \beta^2 \mathbf{W}_2$ ,

$\mathbf{S}_{12} = \mathbf{W}_{12} - \beta \mathbf{W}_2$ ,  $\mathbf{S}_2 = \mathbf{W}_2$  and the benchmark  $BM(\beta, \mathbf{W}) \equiv \sqrt{tr(\mathbf{S}_1)/tr(\mathbf{S}_2)}$ . We write  $\mathcal{S}^{K-1}$  for the  $K - 1$  dimensional unit sphere.

1. For  $e \in \{TSLS, LIML\}$  the Taylor expansion of  $\beta_e^*$  around  $\mu^{-1} = 0$  gives the Nagar (1959) bias

$$N_e(\beta, \mathbf{C}, \mathbf{W}, \boldsymbol{\Omega}) = \mu^{-2} n_e(\beta, \mathbf{C}_0, \mathbf{W}, \boldsymbol{\Omega}) \quad (2.18)$$

with

$$n_{TSLS}(\beta, \mathbf{C}_0, \mathbf{W}, \boldsymbol{\Omega}) = \frac{tr(\mathbf{S}_{12})}{tr(\mathbf{S}_2)} \left[ 1 - 2 \frac{\mathbf{C}_0' \mathbf{S}_{12} \mathbf{C}_0}{tr(\mathbf{S}_{12})} \right] \quad (2.19)$$

$$n_{LIML}(\beta, \mathbf{C}_0, \mathbf{W}, \boldsymbol{\Omega}) = \frac{tr(\mathbf{S}_{12}) - \frac{\sigma_{12}^2}{\sigma_1^2} tr(\mathbf{S}_1) - \mathbf{C}_0' \left( 2\mathbf{S}_{12} - \frac{\sigma_{12}^2}{\sigma_1^2} \mathbf{S}_1 \right) \mathbf{C}_0}{tr(\mathbf{S}_2)} \quad (2.20)$$

2. For  $e \in \{TSLS, LIML\}$ :

$$B_e(\mathbf{W}, \boldsymbol{\Omega}) \equiv \sup_{\beta \in \mathbb{R}, \mathbf{C}_0 \in \mathcal{S}^{K-1}} \left( \frac{|n_e(\beta, \mathbf{C}_0, \mathbf{W}, \boldsymbol{\Omega})|}{BM(\beta, \mathbf{W})} \right) < \infty \quad (2.21)$$

3.  $B_{TSLS}(\mathbf{W}, \boldsymbol{\Omega}) \leq 1$

*Proof.* See the Appendix to this Chapter.. □

**Remark 3.** The Nagar bias is the bias of an approximating distribution. It equals the expectation of the first three terms in the Taylor series expansion of the asymptotic estimator distribution under weak instrument asymptotics. It is therefore always defined and bounded for both TSLS and LIML. While the asymptotic estimator bias may not always exist, our test is still performing well. Under the null hypothesis, the Nagar bias can be large, but under the alternative hypothesis, the Nagar bias is small; see Section 4.2. Under certain conditions, we can also prove that the Nagar bias approximates

the asymptotic estimator bias as the concentration parameter  $\mu^2$  goes to infinity; see Supplementary Materials C.1.

**Remark 4.** We interpret the benchmark  $BM(\beta, \mathbf{W}) = \sqrt{tr(\mathbf{S}_1)/tr(\mathbf{S}_2)}$  as a “worst-case” bias. An ad-hoc approximation of  $\mathbb{E}[\beta_{TSL}^*]$  as a ratio of expectations as in Staiger and Stock (1997) helps convey the intuition:

$$\mathbb{E}[\beta_{TSL}^*] \approx \frac{tr(\mathbf{S}_{12})}{tr(\mathbf{S}_2)[1 + \mu^2]} \quad (2.22)$$

$$\approx \frac{1}{[1 + \mu^2]} \frac{tr(\mathbf{S}_{12})}{\sqrt{tr(\mathbf{S}_2)}\sqrt{tr(\mathbf{S}_1)}} \sqrt{\frac{tr(\mathbf{S}_1)}{tr(\mathbf{S}_2)}} \quad (2.23)$$

The first factor is maximized when instruments are completely uninformative and  $\mu^2=0$ , while the second factor is maximized when first and second stage errors are perfectly correlated (Liu and Neudecker (1995)).

**Remark 5.** In the implementation of our generalized testing procedure, we use the function  $B_e(\mathbf{W}, \mathbf{\Omega})$  to bound the Nagar bias relative to the benchmark. We provide a fast and accurate numerical MATLAB routine for  $B_e(\mathbf{W}, \mathbf{\Omega})$ . For any given value of the structural parameter  $\beta$ , we compute the supremum over  $\mathbf{C}_0 \in \mathcal{S}^{K-1}$  analytically using matrix diagonalization. We then compute the limit of

$$\sup_{\mathbf{C}_0 \in \mathcal{S}^{K-1}} |n_e(\beta, \mathbf{C}_0, \mathbf{W}, \mathbf{\Omega})| / BM(\beta, \mathbf{W})$$

as  $\beta \rightarrow \pm\infty$ . Finally, we numerically maximize the function

$$\sup_{\mathbf{C}_0 \in \mathcal{S}^{K-1}} |n_e(\beta, \mathbf{C}_0, \widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})| / BM(\beta, \mathbf{\Omega})$$

over  $\beta \in [-X, X]$ , where  $X \in \mathbb{R}^+$  is chosen sufficiently large.

## 4.2 Null hypothesis

For a given threshold  $\tau \in [0, 1]$  and matrix  $\mathbf{W} \in \mathbb{R}^{2K \times 2K}$  we define the null and alternative hypotheses for  $e \in \{TSLS, LIML\}$

$$H_e^0 : \mu^2 \in \mathcal{H}_e(\mathbf{W}, \boldsymbol{\Omega}) \quad v.s. \quad H_e^1 : \mu^2 \notin \mathcal{H}_e(\mathbf{W}, \boldsymbol{\Omega}) \quad (2.24)$$

where

$$\mathcal{H}_e(\mathbf{W}, \boldsymbol{\Omega}) \equiv \left\{ \mu^2 \in \mathbb{R}_+ : \sup_{\beta \in \mathbb{R}, \mathbf{C}_0 \in \mathcal{S}^{K-1}} \frac{|N_e(\beta, \mu \sqrt{tr \mathbf{W}_2} \mathbf{C}_0, \mathbf{W}, \boldsymbol{\Omega})|}{BM(\beta, \mathbf{W})} > \tau \right\} \quad (2.25)$$

Under the null hypothesis, the Nagar bias exceeds a fraction  $\tau$  of the benchmark for at least some value of the structural parameter  $\beta$  and some direction of the first stage coefficients  $\mathbf{C}_0$ . On the other hand, under the alternative, the Nagar bias is at most a fraction  $\tau$  of the benchmark for any values  $(\beta, \mathbf{C}_0)$ .

## 4.3 Testing Procedures

We base our test on the statistic  $\widehat{F}_{eff}$ , which is asymptotically distributed as a quadratic form in normal random variables with mean  $1 + \mu^2$ ; see Lemma 1. For a survey of this class of distributions, see Johnson *et al.* (1995, chap. 29). Denote by  $F_{\mathbf{C}, \mathbf{W}_2}^{-1}(\alpha)$  the upper  $\alpha$  quantile of the distribution  $\boldsymbol{\gamma}_2' \boldsymbol{\gamma}_2 / tr(\mathbf{W}_2)$ , where  $\boldsymbol{\gamma}_2 \sim \mathcal{N}_K(\mathbf{C}, \mathbf{W}_2)$  and let

$$c(\alpha, \mathbf{W}_2, x) \equiv \sup_{\mathbf{C} \in \mathbb{R}^K} \{ F_{\mathbf{C}, \mathbf{W}_2}^{-1}(\alpha) \mathbb{1}_{\mathbf{C}' \mathbf{C} / tr(\mathbf{W}_2) < x} \} \quad (2.26)$$

$\mathbb{1}_A(\cdot)$  denotes the indicator function over a set  $A$ . We base the generalized test on the observation that  $\mathcal{H}_e(\mathbf{W}, \boldsymbol{\Omega}) = [0, B_e(\mathbf{W}, \boldsymbol{\Omega})/\tau)$ . The generalized procedure is applicable to both TSLS and LIML, and it rejects the null hypothesis  $H_e^0$  whenever

$$\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\boldsymbol{\Omega}})/\tau) \quad (2.27)$$

**Lemma 2.** *Under Assumptions  $L_\Pi$  and HL the generalized procedure is pointwise*

asymptotically valid, i.e.

$$\sup_{\mathcal{H}_e(\mathbf{W}, \mathbf{\Omega})} \lim_{S \rightarrow \infty} \mathbb{P} \left( \widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau) \right) \leq \alpha$$

Furthermore, provided that  $B(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})$  is bounded in probability

$$\lim_{\mu^2 \rightarrow \infty} \lim_{S \rightarrow \infty} \mathbb{P} \left( \widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau) \right) = 1 \quad (2.28)$$

*Proof.* See the Appendix to this Chapter.. □

The inequality in Theorem 1.3 implies a simplified asymptotically valid test for TSLS, which rejects the null hypothesis  $\mathcal{H}_e(\mathbf{W}, \mathbf{\Omega})$  whenever

$$\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, 1/\tau) \quad (2.29)$$

With  $c(\alpha, \widehat{\mathbf{W}}_2, 1/\tau) \geq c(\alpha, \widehat{\mathbf{W}}_2, B_{TSLS}(\mathbf{W}, \mathbf{\Omega})/\tau)$  the simplified procedure is asymptotically valid and weakly less powerful than the generalized procedure. The simplified test is conservative, in the sense that under the alternative hypothesis, the TSLS Nagar bias is lower than the threshold for any degree of dependence in the second stage.

## 2.5 Computation of Critical Values

We provide two simple methods to compute the critical value  $c(\alpha, \mathbf{W}_2, x)$ . Our first method generates Monte Carlo critical values  $c_m(\alpha, \mathbf{W}_2, x)$ . We obtain estimates of  $F_{\mathbf{C}, \mathbf{W}_2}^{-1}(\alpha)$  as the sample upper  $\alpha$  point from a large number of draws from the distribution of  $\boldsymbol{\gamma}'_2 \boldsymbol{\gamma}_2 / \text{tr}(\widehat{\mathbf{W}}_2)$ , and then maximize over a large set of  $\mathbf{C}$ , such that  $\mathbf{C}'\mathbf{C} / \text{tr}(\mathbf{W}_2) \leq x$ .

The second procedure is based on a curve-fitting methodology first suggested by Patnaik (1949). Patnaik (1949) and Imhof (1961) approximate the critical values of a weighted sum of independent non-central chi-squared distributions by a central  $\chi^2$



with the same first and second moments. We analogously approximate the distribution  $F_{\mathbf{C}, \mathbf{W}_2}$  by a non-central  $\chi^2$  with the same first and second moments. Our approximation errors are therefore bounded by the original Patnaik errors through a triangle inequality. We use

$$F_{\mathbf{C}, \mathbf{W}_2}^{-1}(\alpha) \approx \frac{1}{K_{eff}} F_{\chi_{K_{eff}}^2(K_{eff}\mu^2)}^{-1}(\alpha) \quad (2.30)$$

where  $K_{eff}$  is possibly fractional with

$$K_{eff} = [tr(\mathbf{W}_2)]^2 \frac{1 + 2\mu^2}{tr(\mathbf{W}_2^2) + 2\mathbf{C}'\mathbf{W}_2\mathbf{C}} \quad (2.31)$$

There is a large literature that approximates distributions by choosing a family of distributions and selecting the member that fits best, often by matching lower order moments of the original distribution (Satterthwaite, 1946; Theil and Nagar, 1961; Henshaw, 1966; Pearson, 1959; Grubbs, 1964; Conerly and Mansfield, 1988; Liu *et al.*, 2009). The non-central chi-squared distribution is a natural choice, because it is exact in the homoskedastic case.

While it is hard to assess the accuracy of these curve-fitting approximations analytically, they are often simple and numerically highly accurate (Rothenberg, 1984). Authors demonstrate the degree of accuracy of their approximations using numerical examples. In the Supplementary Materials B.1, we verify that the approximation (2.30) is numerically as accurate as the original central Patnaik distribution for the quadratic forms considered in Imhof (1961); approximation errors are at most 0.7 % points in the important upper 15% tail of the distributions.

Numerical results, such as in Table 2.1, clearly indicate that upper  $\alpha$  quantiles of (2.30) are decreasing in  $K_{eff}$ . Moreover, the upper  $\alpha$  quantile in (2.30) is nondecreasing in the noncentrality parameter  $\mu^2$  (Ghosh, 1973). Taking the supremum over  $\mathbf{C}$  with  $\mathbf{C}'\mathbf{C}/tr(\mathbf{W}_2) < x$ , suggests the Patnaik critical value.

**Definition. (Patnaik Critical Value)** Define the Patnaik critical value as

$$c_P(\alpha, \mathbf{W}_2, x) \equiv F_{(1/K_{eff})\chi_{K_{eff}}^2(xK_{eff})}^{-1}(\alpha) \quad (2.32)$$

with the effective number of degrees of freedom

$$K_{eff} \equiv \frac{tr(\mathbf{W}_2)^2(1+2x)}{tr(\mathbf{W}_2^2) + 2tr(\mathbf{W}_2) \max eval(\mathbf{W}_2)x} \quad (2.33)$$

We numerically analyze the sizes of Monte Carlo and Patnaik critical values for benchmark parameter values  $\alpha = 5\%$  and  $x = 10$ , and find that size distortions are small for both methodologies. Monte Carlo critical values are computed with 40000 draws from  $\gamma_2'\gamma_2/tr(\mathbf{W}_2)$ , and we replace the infinite set of vectors  $\mathbf{C}$  s.t.  $\mathbf{C}'\mathbf{C}/tr(\mathbf{W}_2) < x$  by a finite set of size 500. We use code for  $F_{\mathbf{C}, \mathbf{W}_2}(x)$ .<sup>2</sup> For 400 matrices  $\mathbf{W}_2$  from a diffuse prior with  $K \in \{1, 2, 3, 4, 5\}$  our numerical values for

$\max_{\mathbf{C}'\mathbf{C}/tr\mathbf{W}_2 < x} F_{\mathbf{C}, \mathbf{W}_2}(c_m)$  range between 4.77% and 5.26%, and our numerical values for  $\max_{\mathbf{C}'\mathbf{C}/tr\mathbf{W}_2 < x} F_{\mathbf{C}, \mathbf{W}_2}(c_P)$  range between 5.00% and 5.02%. For further details and MATLAB routines, see Supplementary Materials B.2-B.5.

Our generalized and simplified critical values differ from those proposed in the work of Stock and Yogo (2005) for the TSLS bias, even when first and second stage errors are perfectly conditionally homoskedastic and serially uncorrelated. In this case, the effective F statistic coincides with the Stock and Yogo (2005) test statistic. We obtain different critical values because, unlike them, we use an approximation to evaluate the weak instrument TSLS bias. Moreover, estimating  $\widehat{\mathbf{W}}$  and  $\widehat{\mathbf{\Omega}}$  also generates differences in critical values. The difference between our generalized TSLS critical values and analogous Stock and Yogo (2005) critical values becomes small as the number of instruments becomes large.

---

<sup>2</sup>available at <http://elsa.berkeley.edu/~ruud/cet/pgms.htm> (Imhof, 1961; Koerts and Abrahamse, 1969; Farebrother, 1990; Ruud, 2000)

In the Supplementary Materials B.6, we tabulate Stock and Yogo (2005) 5% critical values for testing the null hypothesis that the TSLS bias exceeds 10% of the OLS bias and our generalized and simplified critical values with a threshold of 10% and size 5%, assuming conditional homoskedasticity and no serial correlation. TSLS critical values are smaller than Stock and Yogo (2005) critical values for  $K = 3, 4$ , but larger than Stock and Yogo (2005) critical values for  $K \geq 5$ . The difference between the TSLS and Stock and Yogo (2005) critical values is always less than 1. The LIML critical values decline more rapidly with the number of instruments than either the TSLS or simplified critical values. The simplified critical values exceed the generalized TSLS critical values, because they use a bound that applies for any form of the matrix  $\mathbf{W}$ .

## 2.6 Empirical Application: Elasticity of Intertemporal Substitution

We now apply our pre-testing procedure to an empirical example, and show that allowing for heteroskedasticity and time series correlation can affect pre-testing conclusions.

The literature has focused on estimating the linearized Euler equation in two standard IV frameworks (Hansen and Singleton, 1983; Campbell and Mankiw, 1989; Hall, 1988; Campbell, 2003).

$$\Delta c_{t+1} = \nu + \psi r_{t+1} + u_{t+1} \text{ and } \mathbb{E}[\mathbf{Z}_{t-1} u_{t+1}] = 0 \quad (2.34)$$

$$r_{t+1} = \xi + (1/\psi)\Delta c_{t+1} + \eta_{t+1} \text{ and } \mathbb{E}[\mathbf{Z}_{t-1} \eta_{t+1}] = 0 \quad (2.35)$$

$\psi$  is the Elasticity of Intertemporal Substitution (EIS),  $\Delta c_{t+1}$  is consumption growth at time  $t + 1$ ,  $r_{t+1}$  is a real asset return, and  $\nu$  is a constant. The vector of instruments is denoted by  $\mathbf{Z}_{t-1}$ . We follow the preferred choice of variables in Yogo (2004), using

as  $r_t$  the real return on the short-term interest rate, and as instruments the nominal interest rate, inflation, consumption growth and the log dividend-price ratio, all lagged twice. We use quarterly data from Yogo (2004).

The EIS determines an agent’s willingness to substitute consumption over time. Its magnitude is important for understanding the dynamics of consumption and asset returns (Epstein and Zin, 1989a, 1991; Campbell, 2003). While time-varying volatility can introduce additional bias into the estimation of the EIS (Bansal and Yaron, 2004), Yogo (2004) argues that under certain types of conditional heteroskedasticity the EIS can still be identified.

Table 2.2 compares pre-tests for weak instruments for 11 countries. Panel A shows weak instrument pre-tests with the ex-post real interest rate as the endogenous variable, while Panel B shows weak instrument pre-tests with consumption growth as the endogenous variable. The non-robust first stage F statistic in column 1 is shown in bold whenever it exceeds the Stock and Yogo (2005) critical value 10.27. This is the 5% critical value for testing the null hypothesis that the TSLS bias exceeds 10% of the OLS bias under the assumption of conditional homoskedasticity and no serial correlation. As in Yogo (2004), this homoskedastic pre-test indicates strong instruments in Panel A, but cannot reject weak instruments in Panel B for almost all countries in the sample.

The second and third columns report the HAC robust first stage F statistic and the effective F statistic computed with a Newey-West kernel and six lags. We show 5% critical values for TSLS, LIML, and simplified pre-tests for the null hypothesis that the respective Nagar bias exceeds 10% of the “worst-case” benchmark.

In Panel A, we see that allowing for heteroskedasticity and serial correlation changes the pre-testing results for some countries, while for other countries all pre-tests yield

**Table 2.2:** *Estimating the Elasticity of Intertemporal Substitution*

**Weak Instrument Pre-Tests**

Panel A: $\Delta c_{t+1} = \nu + \psi r_{t+1} + u_{t+1}$ and $\mathbb{E}[\mathbf{Z}_{t-1}u_{t+1}] = 0$								
Country	$\hat{F}$	$\hat{F}_r$	$\hat{F}_{eff}$	$c_{Simp}$	$c_{TSLS}$	$c_{LIML}$	$\hat{\psi}_{TSLS}$	$\hat{\psi}_{LIML}$
USA	<b>15.53</b>	8.60	7.94	18.20	15.49	9.68	0.06	0.03
AUL	<b>21.81</b>	27.56	17.52	18.36	<b>16.64</b>	<b>10.25</b>	0.05	0.03
CAN	<b>15.37</b>	11.58	12.95	18.95	17.38	<b>11.44</b>	-0.30	-0.34
FR	<b>38.43</b>	41.67	40.29	<b>19.51</b>	<b>17.01</b>	<b>12.89</b>	-0.08	-0.08
GER	<b>17.66</b>	12.47	11.66	18.24	16.30	10.01	-0.42	-0.44
ITA	<b>19.01</b>	25.09	19.44	<b>19.26</b>	<b>17.37</b>	<b>12.98</b>	-0.07	-0.07
JAP	8.64	8.32	5.09	21.66	20.24	18.71	-0.04	-0.05
NTH	<b>12.05</b>	9.31	10.53	18.89	17.18	11.28	-0.15	-0.14
SWD	<b>17.08</b>	28.86	19.82	<b>19.04</b>	<b>15.59</b>	<b>11.65</b>	0.00	0.00
SWT	8.55	6.68	7.19	18.49	15.80	10.38	-0.49	-0.50
UK	<b>17.04</b>	11.78	7.65	20.18	18.72	14.57	0.17	0.16
Panel B: $r_{t+1} = \xi + (1/\psi)\Delta c_{t+1} + \eta_{t+1}$ and $\mathbb{E}[\mathbf{Z}_{t-1}\eta_{t+1}] = 0$								
Country	$\hat{F}$	$\hat{F}_r$	$\hat{F}_{eff}$	$c_{Simp}$	$c_{TSLS}$	$c_{LIML}$	$\hat{\psi}_{TSLS}^{-1}$	$\hat{\psi}_{LIML}^{-1}$
USA	2.93	3.37	2.58	17.61	13.99	10.23	0.68	34.11
AUL	1.79	2.87	2.31	19.89	17.25	15.70	0.50	30.03
CAN	3.03	5.99	2.70	18.19	15.89	9.77	-1.04	-2.98
FR	0.17	0.39	0.22	19.83	18.08	14.09	-3.12	-12.38
GER	0.83	2.48	1.13	18.58	16.98	14.19	-1.05	-2.29
ITA	0.73	0.39	0.47	19.05	16.96	11.63	-3.34	-14.81
JAP	1.18	2.17	2.00	17.94	13.93	15.58	-0.18	-21.56
NTH	0.89	3.62	1.84	19.00	16.13	15.30	-0.53	-6.94
SWD	0.48	0.81	0.83	17.24	12.51	9.73	-0.10	-399.86
SWT	0.97	2.28	1.56	20.21	18.76	16.47	-1.56	-2.00
UK	2.52	3.95	2.55	17.94	15.64	14.50	1.06	6.21

NOTE:  $\Delta c$  is consumption growth and  $r$  is the ex-post real short-term interest rate. We instrument using twice lagged nominal interest rate, inflation, dividend-price ratio, and consumption growth. HAC variance-covariance matrix  $\widehat{\mathbf{W}}$  estimated with OLS and Newey-West kernel with six lags. F statistic in bold when it exceeds the critical value of 10.27. This is the 5% critical value for testing the null hypothesis that the TSLS bias exceeds 10% of the OLS bias under the assumption of conditional homoskedasticity and no serial correlation (Stock and Yogo, 2005). We show simplified, TSLS, and LIML critical values  $c_{Simp} = c_P(5\%, \widehat{\mathbf{W}}_2, 10)$ ,  $c_{TSLS} = c_P(5\%, \widehat{\mathbf{W}}_2, 10 \times B_{TSLS}(\widehat{\mathbf{W}}, \widehat{\boldsymbol{\Omega}}))$ , and  $c_{LIML} = c_P(5\%, \widehat{\mathbf{W}}_2, 10 \times B_{LIML}(\widehat{\mathbf{W}}, \widehat{\boldsymbol{\Omega}}))$ . Critical values are in bold when exceeded by  $\hat{F}_{eff}$ .  $\hat{\psi}_{TSLS}$ ,  $\hat{\psi}_{LIML}$ ,  $(1/\psi)_{TSLS}$  and  $(1/\psi)_{LIML}$  are TSLS and LIML estimates of the EIS and its inverse.

the same conclusion. The effective F statistic can be smaller or larger than the regular or robust F statistics. Simplified critical values always exceed TSLS critical values. LIML critical values tend to be smallest.

The results in Table 2.2A for the U.S. are particularly striking. While the U.S. regular F statistic clearly exceeds the homoskedastic threshold of 10.27, the robust and effective F statistics are significantly smaller. The effective F does not exceed the simplified, TSLS, or LIML critical values, so we cannot reject the null hypothesis of weak instruments under heteroskedasticity and serial correlation.

Panel B shows weak instrument pre-tests for the instrumental variable estimation of the inverse of the EIS. For this estimation, the results are consistent between homoskedastic and HAC weak instrument pre-tests. We cannot reject that instruments are weak for any of the countries in the sample.

The last two columns in Table 2.2 show the point estimates for  $\psi$  and  $1/\psi$ . For those cases where we can reject weak instruments under heteroskedasticity and serial correlation, the corresponding EIS point estimates are close to zero and often negative. Additional caution is, however, warranted in this interpretation, because as the number of countries increases, we are more and more likely to reject weak instruments at least once.

Our results confirm Yogo (2004)'s finding that the EIS is small and close to zero. However, we also note that conditional heteroskedasticity and serial correlation may further weaken instruments and may affect TSLS and LIML bias in several of the country-specific regressions.

## 2.7 Conclusion

Heteroskedasticity, serial correlation, and panel data clustering can affect instrument strength. This paper develops a robust test for weak instruments that allows empirical researchers to test the null hypothesis that the TSLS or LIML Nagar bias is large relative to a benchmark.

The test is based on a scaled version of the regular F statistic. Critical values depend on the covariance matrix of the reduced form coefficients and errors. Our general test requires computational work to evaluate the Nagar bias of TSLS or LIML. A simplified conservative version does not require this step, but is only available for TSLS. Critical values can then be implemented as quantiles of a non-central chi-squared distribution with non-integer degrees of freedom.

Pre-tests based on the robust (or non-robust) first stage F statistic with Stock and Yogo (2005) critical values are commonly applied outside the conditionally homoskedastic serially uncorrelated framework. However, to the best of our knowledge, there is no analysis supporting this practice. This paper offers an alternative: a simple, asymptotically valid test that should be used for conditionally heteroskedastic, time series, and clustered panel data.

# Chapter 3

## Efficient Conditionally

## Similar-on-the-Boundary Tests <sup>1</sup>

### 3.1 Introduction

THERE IS A STANDARD APPROACH IN ECONOMETRICS to test statistical hypotheses in the presence of nuisance parameters. First, one finds a point estimate for the parameter of interest. Second, one approximates the distribution of the estimator's sampling error. Third, one estimates the relevant nuisance parameters. The standard test is implemented by comparing the estimator's *null* sampling error (i.e.,  $\hat{\theta} - \theta_0$ ) with the quantiles of the estimated distribution. Despite its prevalence, there is now a large body of work—both empirical and theoretical—documenting problems with this practice in the context of several widely used models. Three important examples are: *Linear Instrumental Variables Regression* (IV) [Nelson and Startz (1990), Bound *et al.* (1995), Staiger and Stock (1997), Stock, Wright, and Yogo (2002)]; the *Generalized Method of Moments* (GMM) [Andersen and Sørensen (1996), Hansen *et al.* (1996),

---

<sup>1</sup>Job-Market Paper



Stock and Wright (2000), Mavroeidis, Plagborg-Møller, and Stock (2012)]; and *Structural Vector Autoregressions identified using external instruments* (SVAR) [Montiel Olea, Stock, and Watson (2012)]. The main practical concern is that the actual rate of Type I error for standard testing procedures (e.g., Wald tests) can be very different from the nominal target and dramatically changes with the values of the nuisance parameters.

This paper studies point and one-sided testing problems in IV, GMM, and SVARs from a different perspective. I analyze a general class of parametric hypothesis testing problems with a key characteristic: a *boundary-sufficient statistic*. Broadly speaking, a statistic  $X_2$  is boundary sufficient if any movement along the set of null parameter values that are the closest to the alternative hypothesis—i.e., the boundary of the null—affects the distribution of the data  $(X_1, X_2)$  only through its effect on  $X_2$ . I show that this property arises naturally in the limiting experiments associated with the three aforementioned examples.<sup>2</sup> In each of these cases, it is possible to control the rate of Type I error by conditioning on the corresponding boundary-sufficient statistic: an adjusted vector of Ordinary Least Squares reduced-form first-stage coefficients in the IV model, an adjusted derivative of the sample moment condition in GMM, and a linear transformation of the sample covariance between the reduced form errors in the vector autoregression and the external instruments used to identify the SVAR.<sup>3</sup>

The main theoretical contribution of this paper is a new class of tests for hypothesis testing problems with a boundary-sufficient statistic. The *Efficient Conditionally Similar-on-the-Boundary tests* (henceforth, ECS tests) are minimizers of a weighted sum

---

<sup>2</sup>I use the phrase *limiting experiment* in the modern sense of Müller (2011) and not in the classical sense of Le Cam (1986). Thus, a limiting experiment refers to a statistical model derived from a set of weak convergence assumptions. Section 3.5 presents a detailed explanation of the concept.

<sup>3</sup>The remarkable paper of Moreira (2003) introduced the idea of “conditioning” as a device to control the rate of Type I error in Structural Equations Models.

of the average rates of Type I and Type II error subject to a *conditional similarity-on-the-boundary constraint*. Their main claim for optimality, albeit decision-theoretic, is of a very applied nature: there is no other test—among those that condition the accept/reject decision on the realizations of a boundary-sufficient statistic—with smaller rates of Type I and Type II error. That is, ECS tests are *admissible* within the class of conditionally similar-on-the-boundary tests. Neither Moreira’s (2003) Conditional Likelihood Ratio (CLR) for testing a point hypothesis in IV nor Kleibergen’s (2007) extensions of the CLR to GMM have been shown to satisfy this property.

The admissibility result can be further strengthened. This paper shows that ECS tests are admissible within the class of *all* tests, provided the boundary-sufficient statistic is *boundedly complete* (as defined by Lehmann and Romano (2005)) and the rates of Type I and Type II error vary continuously over the parameter space. These assumptions are satisfied in several IV, GMM, and SVAR settings. The result is relevant for applied econometrics. For instance, neither the Two-Stage Least Squares (TSLS) nor the Limited Information Maximum Likelihood (LIML) Wald tests are known to be admissible, not even in the context of a Gaussian, independent, homoskedastic model.<sup>4</sup> Hence, even if practitioners do not regard similarity-on-the-boundary as a desirable property—which the widespread use of the TSLS Wald test in IV regression suggests—there is still a strong justification to use ECS procedures, for it is not possible to find a non-similar test with better rates of Type I and Type II error.

The theory developed in this paper provides new insights about hypotheses testing in IV, GMM, and SVARs. There are five main results with an emphasis on point

---

<sup>4</sup>Consider the linear IV regression model with a single endogenous regressor ( $\beta$ ) under the following assumptions. Suppose that the instruments are non-stochastic (fixed) and suppose that the reduced-form errors are independent and identically distributed as a bivariate Gaussian random vector with known covariance matrix. To the best of my knowledge, there are no finite-sample optimality claims available for either the TSLS or the LIML Wald tests. In other words, there is no theoretical support for the use of the test that rejects  $\mathbf{H}_0 : \beta = \beta_0$  in favor of  $\mathbf{H}_1 : \beta \neq \beta_0$  for large values of  $(\hat{\beta}_{\text{TSLS}} - \beta_0)^2 / (\widehat{\text{var}}(\hat{\beta}_{\text{TSLS}} - \beta_0))$ .

testing—in which case, ECS tests are simply maximizers of weighted average power (for a full-support *prior*) subject to a conditional similarity constraint. First, I show that the Anderson and Rubin (1949) test (henceforth, AR) is ECS in just-identified IV models with Gaussian reduced-form errors, independent observations, fixed instruments, and an arbitrary number of endogenous regressors. Furthermore, a robust version of the AR test is shown to be ECS in the limiting experiment of *weakly just-identified* IV models with heteroskedastic, autocorrelated, and/or clustered data. The priors over the structural parameters of the IV model (denoted  $\beta$  and  $\Pi$ ) for which the AR test maximizes weighted average power have an interesting property: there are no other priors for which the implied distribution over the reduced-form parameters ( $\Pi\beta$  and  $\Pi$ ) is Gaussian, centered at zero, and with the same covariance matrix as the distribution of their sample counterparts.

Second, I derive new ECS tests—for point and one-sided null hypotheses—in the over-identified IV model studied by Andrews *et al.* (2006) and Chamberlain (2007). The ECS test for the point hypothesis problem enjoys basic optimality properties that neither CLR nor the TSLS (LIML) Wald tests have been shown to satisfy. The “conditional” critical region of the new test—which can be expressed in terms of the AR and the Lagrange Multiplier (LM) statistics—admits a simple interpretation: if the LM is below (above) its conventional  $\chi_1^2$  critical value, the ECS test automatically adjusts upwards (downwards) the  $\chi_k^2$  threshold for the AR. The magnitude of the adjustment depends on the value of the boundary-sufficient statistic and the ECS test rejects the null hypothesis whenever the AR exceeds the adjusted critical value. This procedure is also ECS in models in which the reduced-form ordinary least-squares coefficients exhibit a “Kronecker” asymptotic covariance matrix, for example, the proportional heteroskedasticity/autocorrelation models used in Montiel Olea and Pflueger (2012).

Third, I derive a limiting experiment for GMM models with one scalar parameter

and  $m$  moment conditions. The statistical experiment is derived by considering a set of Gaussian weak convergence assumptions for both the sample moment condition and its derivative. I provide a set of sufficient conditions under which the GMM  $S$ -test of Stock and Wright (2000) is ECS in the limiting experiment.

Fourth, I present general ECS tests for over-identified GMM models in which the strength of identification is controlled by a finite-dimensional nuisance parameter. The tests are specialized to non-homoskedastic and/or serially correlated weakly identified IV with one endogenous regressor. In this context, the implementation of the ECS test requires two numerical exercises. First, numerical integration is required to compute an *integrated likelihood* in the ECS test statistic. Second, Monte-Carlo methods are used to compute the quantiles of the empirical distribution of the ECS test statistic, conditional on the boundary-sufficient statistic.

Finally, I derive ECS tests for the limiting experiment of SVARs identified by external instruments, as defined in Montiel Olea *et al.* (2012). The external instruments are random variables correlated with a target shock  $i$ , uncorrelated with the other structural shocks in the model, and excluded from the vector autoregression. The object of interest is the dynamic effect of the structural shock  $i$  over variable  $j$  at horizon  $h$ . The fifth result in this paper shows that the test used by Montiel, Stock, and Watson (2012) to build confidence intervals for dynamic effects is ECS—provided there is only one external instrument for the target shock  $i$ . The ECS test rejects for large values of the sample covariance between the instrument and a linear combination of the reduced-form shocks in the vector autoregression. This paper also presents an ECS test for the over-identified SVAR model.

The remainder of this paper is organized as follows. Section 3.2 presents the basic elements of a parametric testing problem (sample space, parameter space, statistical model, test, Type I/II error, risk, and admissibility) and the main regularity assump-

tions (which I denote TC1, TC2, C). Section 3.3 defines *boundary sufficiency*, which is the key concept in this paper. Throughout both sections, a Gaussian “quasi-shift” model is used to illustrate the main concepts and assumptions. Section 3.4 presents the ECS tests, their main theoretical properties, and the main result concerning their implementation. Section 3.5 derives ECS tests for each of the examples discussed in the introduction. Section 3.6 presents a summary of the main results and concludes. All proofs are collected in the Appendix.

## 3.2 Basic Definitions and Assumptions

Section 3.2.1 presents the three basic elements of a parametric testing problem: sample space, parameter space, and statistical model. This section also defines the *boundary of the null hypothesis* ( $\text{Bd}\Theta_0$ ) and presents Assumptions TC1 and TC2, both of which impose restrictions on the types of null hypotheses under consideration. A simple example (Gaussian quasi-shift model) is used to illustrate the concepts and assumptions.

Section 3.2.2 defines the rates of Type I/Type II error of a test  $\phi$ , both of which are summarized by the risk function,  $R(\phi, \theta)$ . Just as in classical decision theory, risk is used to define the optimality criterion for test selection: *admissibility*. Section 3.2.2 also introduces Assumption C, which imposes a continuity restriction on the rates of Type I and Type II error.

The main definitions in this section follow Chamberlain (2007); Chapters 2 and 5 in Ferguson (1967); and Chapter 4 in Linnik (1968).

NOTATION PRELIMINARIES: Let  $\mathcal{B}(\mathbb{R}^n)$  denote the Borel  $\sigma$ -algebra on  $\mathbb{R}^n$ . For any set  $\mathcal{S} \in \mathcal{B}(\mathbb{R}^n)$ , let  $\mathcal{B}(\mathbb{R}^n)_{\mathcal{S}}$  denote the sub-space  $\sigma$ -algebra.<sup>5</sup> *Measurability* of the

---

<sup>5</sup>That is,  $\mathcal{B}(\mathbb{R}^n)_{\mathcal{S}} \equiv \{\mathcal{S} \cap F \mid F \in \mathcal{B}(\mathbb{R}^n)\}$ .

function  $f : \mathcal{S} \rightarrow \mathbb{R}$  is always relative to the measurable spaces  $(\mathcal{S}, \mathcal{B}(\mathbb{R}^n)_{\mathcal{S}})$ - $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . The integral of  $f$  with respect to the lebesgue measure in  $\mathbb{R}^n$  is denoted by  $\int_{\mathcal{S}} f(s) ds$ . Integration with respect to a different measure  $\mu$  is denoted  $\int_{\mathcal{S}} f(s) d\mu(s)$ . All the  $\mathbb{R}^m$ -valued random variables in this paper are assumed to be absolutely continuous with respect to the lebesgue measure in  $\mathbb{R}^m$ , unless otherwise noted. Thus, random variables with discrete support are ruled out.

### 3.2.1 Basic Elements of a Parametric Testing Problem

**SAMPLE SPACE, PARAMETER SPACE, AND STATISTICAL MODEL:** There is a random variable  $X$  that takes values in the *sample space*  $\mathbf{X} \subseteq \mathbb{R}^s$ . There is a *parameter space*  $\Theta \subseteq \mathbb{R}^p$  whose elements  $\theta \in \Theta$  are used to index a set of probability density functions over the sample space,  $X \sim f(x, \theta)$ . The collection  $\{f(\cdot, \theta)\}_{\theta \in \Theta}$  is called a *statistical model*. The mapping  $f : \mathbf{X} \times \Theta \rightarrow \mathbb{R}_+$  is called the *likelihood function*. It is assumed that the sample space has a product space structure. Consequently,  $X$  can be written as a random vector  $(X_1, X_2)$  with realizations  $(x_1, x_2) \in \mathbb{R}^{s_1} \times \mathbb{R}^{s_2}$ ,  $s_1 + s_2 = s$ .

**NULL HYPOTHESIS:** Let  $\Theta_0$  be a strict subset of the parameter space. There is a null hypothesis  $\mathbf{H}_0$  that states  $X \sim f(x, \theta)$  for some  $\theta \in \Theta_0$ . The hypothesis testing problem is abbreviated  $\mathbf{H}_0 : \theta \in \Theta_0$  vs.  $\mathbf{H}_1 : \theta \in \Theta_1 \equiv \Theta \setminus \Theta_0$ , and it is denoted by the tuple  $(\mathbf{X}, \Theta, f, \Theta_0)$ .

**BOUNDARY OF THE NULL HYPOTHESIS :** The set  $\text{Bd}\Theta_0$  plays an important role in this paper. For the sake of formality, I present a general topological definition of this set. Let  $\mathcal{T}$  be the subspace topology on  $\Theta \subseteq \mathbb{R}^p$  and let  $\tau_\theta$  denote an open neighborhood

of  $\theta \in \Theta$ ; i.e.,  $\theta \in \tau_\theta$  and  $\tau_\theta \in \mathcal{T}$ . Define

$$\text{Bd}\Theta_0 \equiv \{\theta \in \Theta \mid \tau_\theta \cap \Theta_0 \neq \emptyset \text{ and } \tau_\theta \cap \Theta_1 \neq \emptyset, \forall \tau_\theta \in \mathcal{T}\}.$$
<sup>6</sup>

Intuitively, the boundary of the *null set*  $\Theta_0$  contains those elements of the null that are the closest to the alternative.<sup>7</sup>

ASSUMPTIONS CONCERNING THE STRUCTURE OF THE NULL: All the hypotheses testing problems considered in this paper satisfy the following assumptions:

**ASSUMPTION TC1:**  $\#(\text{Bd}\Theta_0) > 1$ .<sup>8</sup>

**ASSUMPTION TC2:**  $\Theta_0$  is closed relative to  $(\Theta, \mathcal{T})$ .

Assumptions TC1 and TC2 imply that  $\Theta_0$  is composite:  $\text{Bd}\Theta_0 \subseteq \Theta_0$  and therefore the null set is not a singleton.<sup>9</sup> Not all hypothesis testing problems with a composite null satisfy Assumption TC1. For instance, in a one-dimensional Gaussian location model with parameter  $\mu$ , the hypothesis  $\mathbf{H}_0 : \mu \leq 0$  is closed and composite. However, the boundary of the null contains only one point:  $\mu = 0$ . The main property used in this paper, *boundary sufficiency*, is only defined for models in which  $\text{Bd}\Theta_0$  has more than one element.

---

<sup>6</sup>The topological boundary of  $\mathcal{A} \subseteq \Theta$  is usually defined as the intersection of two sets: the closure of  $\mathcal{A}$  and the closure of  $\Theta \setminus \mathcal{A}$ ; see Munkres (2000) pp. 95, 102 (Exercise 19). The definition presented here is based on the characterization of closure provided in Munkres (2000), Theorem 17.5a, p. 96.

<sup>7</sup>If  $\theta$  belongs to the boundary, any open ball  $\tau_\theta$  contains an element of the alternative hypothesis. In this sense, there is always a “nearby” element of  $\Theta_1$ . If, however,  $\theta$  belongs to  $\Theta_0 \setminus \text{Bd}\Theta_0$ , then the latter statement no longer holds: there is a neighborhood of  $\theta_0$  that does not contain elements of  $\Theta_1$ .

<sup>8</sup> $\#A$  is defined as the cardinality of the set  $A$ .

<sup>9</sup>TC should be read as *topologically composite*.

EXAMPLE—GAUSSIAN QUASI-SHIFT MODEL: This parametric testing problem—which is intrinsically connected with a just-identified instrumental variable regression (See Section 3.5.1)—illustrates the concepts discussed thus far. Let the sample space  $\mathbf{X} \equiv (\mathbf{X}_1, \mathbf{X}_2) = \mathbb{R}^n \times \mathbb{R}^{n^2}$ ,  $n \in \mathbb{N}$ . Let  $\boldsymbol{\mu}_1$  be an  $n \times 1$  vector and let  $\boldsymbol{\mu}_2 = [\mu_{21}, \mu_{22}, \dots, \mu_{2n}]$  be a  $n \times n$  matrix, not necessarily of full rank. Let the parameter space be given by

$$\{\text{vec}(\boldsymbol{\mu}_1, \mu_{21} \dots \mu_{2n}) : \boldsymbol{\mu}_1 \in \mathbb{R}^n \text{ and } \mu_{2i} \in \mathbb{R}^n \forall i = 1 \dots n\}.$$

Consider the statistical model:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{n+n^2} \left( \begin{array}{c} \boldsymbol{\mu}_2 \boldsymbol{\mu}_1 \\ \text{vec}(\boldsymbol{\mu}_2) \end{array}, \begin{array}{c} \mathbb{I}_{n+n^2} \end{array} \right)$$

and the testing problems:

$$\mathbf{H}_0 : \boldsymbol{\mu}_1 = \mathbf{0} \quad \text{vs.} \quad \mathbf{H}_1 : \boldsymbol{\mu}_1 \neq \mathbf{0} \quad (\text{“Point-null”})$$

or

$$\mathbf{H}_0 : \boldsymbol{\mu}_1 \leq \mathbf{0} \quad \text{vs.} \quad \mathbf{H}_1 : \boldsymbol{\mu}_1 \not\leq \mathbf{0} \quad (\text{“One-sided”}).$$

*Boundary of the null in the Gaussian Quasi-shift Model:* In the point-null testing problem  $\text{Bd}\Theta_0 = \Theta_0$ , the boundary of the null *is the null hypothesis itself*. In the one-sided problem, the set  $\text{Bd}\Theta_0$  contains the set of parameter values  $\boldsymbol{\mu}_1 \leq \mathbf{0}$  for which at least one of the components is equal to zero.

*Assumption TC1:* The parameter  $\boldsymbol{\mu}_2$  is a nuisance parameter on  $\text{Bd}\Theta_0$ . Therefore, neither of the testing problems considered above have a set  $\text{Bd}\Theta_0$  with only one element. Therefore, Assumption TC1 is verified.



*Assumption TC2:* The null set in both testing problems is a closed set relative to the standard topology in  $\mathbb{R}^{n+n^2}$ .

### 3.2.2 Tests, Type I/Type II Error and Risk Function

TESTS: A *test* is a measurable mapping

$$\phi : \mathbf{X} \rightarrow [0, 1],$$

where the scalar  $\phi(x)$  is interpreted as the probability of rejecting  $\mathbf{H}_0$  (in favor of  $\mathbf{H}_1$ ) after a realization  $x$  of  $X$ . Therefore, a test is a summary of the decision of whether to accept or reject  $\mathbf{H}_0$  for all *data sets*,  $x$ , in the sample space. Let  $\mathcal{C}$  denote the class of all tests.

TYPE I AND TYPE II ERROR: Fix a test  $\phi$ . The *rate of Type I error* of test  $\phi$  at  $\theta \in \Theta_0$  is defined as

$$\mathbb{E}_\theta[\phi(X)] = \int_{\mathbf{X}} \phi(x) f(x, \theta) dx.$$

This rate refers to the probability of rejecting the null hypothesis when the true parameter belongs to the null set. Likewise, the *rate of Type II error* of  $\phi$  at  $\theta \in \Theta_1$  is defined as

$$1 - \mathbb{E}_\theta[\phi(X)] = 1 - \int_{\mathbf{X}} \phi(x) f(x, \theta) dx.$$

When both  $\mathbf{H}_0$  and  $\mathbf{H}_1$  are composite, the Type I and Type II errors vary over  $\Theta_0$  and  $\Theta_1$ . These variations are summarized by the risk function, defined as

$$R(\phi, \theta) \equiv \begin{cases} \mathbb{E}_\theta[\phi(X)] & \text{if } \theta \in \Theta_0 \\ 1 - \mathbb{E}_\theta[\phi(X)] & \text{if } \theta \in \Theta_1. \end{cases}$$

ADMISSIBILITY: The optimality criterion used in this paper is that of admissibility. This classical decision theoretic concept provides a natural ordering over tests based on the risk function. Let  $\mathcal{C}^* \subseteq \mathcal{C}$  be a class of tests that contain  $\phi$ . Let  $\phi'$  be an arbitrary element of  $\mathcal{C}^*$ .

**DEFINITION 1:** (Ferguson (1967), p. 54) The test  $\phi$  is **admissible** within the class  $\mathcal{C}^*$  if there is no  $\phi' \in \mathcal{C}^*$  such that  $R(\phi', \theta) \leq R(\phi, \theta)$  for all  $\theta \in \Theta$ , with strict inequality for at least one  $\theta \in \Theta$ .<sup>10</sup>

Tests that are inadmissible within a class  $\mathcal{C}^*$  can be improved (that is, smaller rates of Type I and Type II error can be achieved) all over the parameter space. Thus, admissibility is a minimal requirement that a test must satisfy.

ASSUMPTIONS ON THE BEHAVIOR OF THE RISK FUNCTION: The behavior of  $R(\phi, \theta)$  is restricted by imposing a regularity assumption on the statistical models under study:

**ASSUMPTION C:** For any measurable set  $\mathcal{F} \in \mathcal{B}(\mathbb{R}^s)_{\mathbf{X}}$ , the real-valued function  $P_{\mathcal{F}}(\theta) \equiv \int_{\mathcal{F}} f(x, \theta) dx$  is continuous in  $\theta$ , for every  $\theta \in \Theta$ .

Assumption C implies that for any test  $\phi$ ,  $\mathbb{E}_{\theta}[\phi(X)]$  is a continuous function of  $\theta$ . Therefore, this paper only considers problems in which the risk function of any test is continuous on both  $\text{Int}(\Theta_0)$  and  $\text{Int}(\Theta_1)$ .<sup>11</sup> A sufficient condition for Assumption C is

---

<sup>10</sup>Define an “ordering” over tests as a binary relation  $\succ$  in the space of all tests that verifies two properties. The first one is asymmetry:  $\phi \succ \phi' \implies \phi' \not\succ \phi$ . The second one is transitivity:  $\phi \succ \phi'$  and  $\phi' \succ \phi''$  implies  $\phi \succ \phi''$ . Admissibility induces an ordering through the “weakly dominated” binary relation: a test  $\phi'$  *weakly dominates*  $\phi$  if  $R(\phi', \theta) \leq R(\phi, \theta)$  with strict inequality for at least one  $\theta \in \Theta$ .

<sup>11</sup> $\text{Int}(\Theta_i)$ ,  $i = \{0, 1\}$ , denotes the topological interior of the set  $\Theta_i$ . That is,  $\text{Int}(\Theta_i) \equiv \Theta_i \setminus \text{Bd}\Theta_i$ .

the continuity (in  $\theta$ ) of  $f(x, \theta)$ , for each  $x \in \mathbf{X}$ . See Lemma 5.1 in Wald (1950), p. 133; or Theorem 10 in Berger (1985), p. 545.

### 3.3 Testing problems with a boundary-sufficient statistic

This paper focuses on the study of testing problems with a *boundary-sufficient statistic*. This statistical property is common to Linear IV, weakly identified GMM, Structural VARs, and some other problems with nuisance parameters; for example, the Linear Regression Model with a sign restriction in Elliott *et al.* (2012) and the predictive regression model with nearly integrated regressors studied in Stock and Watson (1996), Jansson and Moreira (2006), and Elliott *et al.* (2012).

This section introduces the notion of a boundary-sufficient statistic and a boundary conditional likelihood. These concepts are further illustrated using the Gaussian quasi-shift experiment.

**BOUNDARY SUFFICIENCY:** Boundary sufficiency is intuitively described as follows. Let  $f(x_1, x_2, \theta)$  be a statistical model for the elements of the product sample space  $\mathbf{X}$ . The statistic  $X_2$  is boundary sufficient if movements of  $\theta$  along the boundary of  $\Theta_0$  affect the distribution of the data  $(X_1, X_2)$  *only* through its effect on  $X_2$ . Formally, this is captured by requiring the likelihood to satisfy the following decomposition.

**DEFINITION 2:** The statistic  $X_2$  is *boundary sufficient* for the testing problem  $(\mathbf{X}, \Theta, f, \Theta_0)$  if

$$f(x_1, x_2, \theta) = g(x_1, x_2)h(x_2, \theta) \quad \text{for every } \theta \in \text{Bd } \Theta_0,$$

where  $g(\cdot, x_2)$  is a probability density function with support given by the set  $\mathbf{X}_1(x_2) \equiv \{x_1 \in \mathbf{X}_1 | (x_1, x_2) \in \mathbf{X}\}$  and  $\{h(x_2, \theta)\}_{\theta \in \text{Bd}\Theta_0}$  is a statistical model for the random variable  $X_2$ .

**Remark 6.** Section 3.5 in this paper shows that in the IV model the boundary-sufficient statistic relates to the OLS estimate of the first-stage coefficient; in GMM,  $X_2$  is a function of the derivative of the sample moment condition; in SVARs it is a function of the correlations of the reduced-form VAR errors and external instruments used to identify the structural shocks.

BOUNDARY CONDITIONAL LIKELIHOOD: In general,  $g(x_1, x_2)$  corresponds to the density of the conditional distribution of  $X_1$  given  $X_2$ , which does not depend on the element  $\theta \in \text{Bd}\Theta_0$  at which the likelihood is evaluated. In light of this observation,  $g(x_1, x_2)$  is denoted as  $f_{\text{Bd}}(x_1 | x_2)$ ; and it is called the *boundary conditional likelihood*.<sup>12</sup>

BOUNDARY SUFFICIENCY IN THE GAUSSIAN QUASI-SHIFT MODEL: For simplicity, consider the point-null problem  $\mathbf{H}_0 : \boldsymbol{\mu}_1 = 0$ . The Gaussian quasi-shift model evaluated at the boundary of null ( $\boldsymbol{\mu}_1 = 0$ ) becomes

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{n+n^2} \begin{pmatrix} \mathbf{0} \\ \text{vec}(\boldsymbol{\mu}_2) \end{pmatrix}, \quad \mathbb{I}_{n+n^2}.$$

Consequently, any movement along  $\text{Bd}\Theta_0$ —which corresponds to a change in  $\boldsymbol{\mu}_2$  while keeping  $\boldsymbol{\mu}_1 = 0$ —affects  $(X_1, X_2)$  only through its effect in the location parameter of  $X_2$ . Hence,  $X_2$  is a boundary-sufficient statistic. The boundary conditional likelihood

---

<sup>12</sup>Note that for any  $\theta \in \text{Bd}\Theta_0$ :

$$f(x_1, x_2; \theta) / \int_{\mathbf{X}_1(x_2)} f(x_1, x_2; \theta) dx_1 = f(x_1, x_2; \theta) / h(x_2; \theta) = g(x_1, x_2).$$

is given by the density of  $X_1 \sim \mathcal{N}_n(\mathbf{0}, \mathbb{I}_n)$ .

The theory developed in this paper is general enough to include “point-null” and some “one-sided” hypothesis testing problems with nuisance parameters. For instance, when  $\mathbf{H}_0 : \boldsymbol{\mu}_1 \leq 0$  and  $n = 1$ , the statistical model satisfies boundary sufficiency.

### 3.4 Main Results

This paper provides a systematic approach to generate admissible tests within the class of *conditionally similar-on-the-boundary tests*; that is, testing procedures that control the rate of Type I error on the boundary of the null hypothesis by conditioning the accept/reject decision on the realizations of a boundary-sufficient statistic,  $X_2$  (see Definition 3 below). The new tests derived are shown to be admissible within the class of all tests whenever  $X_2$  is *boundedly complete* (see Definition 5 below). The latter property is verified in IV and also in some GMM, SVARs models).

This section starts by presenting the class of *Efficient Conditionally Similar Tests* (subsequently abbreviated ECS tests), which are defined as minimizers of average risk in the class of tests that are *conditionally similar on the boundary*. The main results of this section are Theorems 1 and 2.

Theorem 1a shows that the optimization problem defining ECS tests has a solution. Theorem 1b shows that the ECS tests are admissible within the class of all procedures that control Type I error by means of a boundary-sufficient statistic. Theorem 1c extends the admissibility result to the class of all tests, provided the boundary-sufficient statistic is also boundedly complete. Theorem 2 provides the basis for the implementation of ECS tests. Under certain regularity conditions, the new tests are implemented by comparing

- a) The ratio of “weighted difference of integrated likelihoods” relative to the bound-

ary conditional likelihood, against;

b) A *critical value function* that depends on the boundary-sufficient statistic.

### 3.4.1 ECS Tests

Let  $X_2$  be a boundary-sufficient statistic and let  $h(x_2, \theta)$  denote the probability density function of  $X_2$  parameterized by  $\theta \in \text{Bd}\Theta_0$ .

**DEFINITION 3:** A test  $\phi$  is  $\alpha$ -*conditionally similar on the boundary of the null* (abbreviated  $\alpha$ -csb) if

$$\mathbb{E}_\theta[\phi(X_1, X_2) \mid X_2] = \alpha$$

for all  $\theta \in \text{Bd}\Theta_0$ , and for all  $x_2 \in \mathbf{X}_2$  except perhaps in a set having probability zero under all distributions  $\{h(x_2, \theta)\}_{\theta \in \text{Bd}\Theta_0}$ .

Let  $\mathcal{C}_{X_2}(\alpha\text{-csb})$  denote the class of all  $\alpha$ -csb tests. The law of iterated expectations implies that an  $\alpha$ -csb test is  $\alpha$ -*similar on the boundary* ( $\alpha$ -sb), this is:

$$\mathbb{E}_\theta[\phi(X)] = \alpha, \quad \forall \theta \in \text{Bd}\Theta_0.$$

Similarity and conditional similarity are classical concepts in statistical decision theory.<sup>13</sup> However, to the best of my knowledge, there are no general results concerning the construction of admissible similar or conditionally similar tests in the presence of a boundary-sufficient statistic.

---

<sup>13</sup>Similarity was first introduced by Neyman (1935) and it has been extensively studied by Linnik (1968). Neyman does not use the word “*similarity*” in his paper. Instead he refers to a critical region whose area is well-determined by the (composite) hypothesis to verify (*ensemble critique d’aire ‘ $\alpha$ ’ bien déterminée par l’hypothèse à vérifier*). Linnik (1968) refers to such regions as  $\alpha$ -similar regions.

EFFICIENT CONDITIONALLY SIMILAR TESTS: Let  $p_i(\theta)$  denote a full-support probability density function over  $\text{Int}\Theta_i$ , for  $i = \{0, 1\}$  and let  $\tau \in (0, 1)$ .<sup>14</sup>

**DEFINITION 4:** A test  $\phi^*$  is  $\alpha$ -**Efficient Conditionally Similar on the Boundary** if

$$\phi^* \in M(\tau, p_1, p_0)$$

where

$$M(\tau, p_1, p_0) \equiv \arg \min_{\phi \in \mathcal{C}_{X_2}(\alpha\text{-csb})} \tau \int_{\text{Int } \Theta_1} R(\phi, \theta) p_1(\theta) d\theta + (1 - \tau) \int_{\text{Int } \Theta_0} R(\phi, \theta) p_0(\theta) d\theta. \quad (3.1)$$

ECS tests are built in the following way: a full-support prior on the interior of the alternative set  $\Theta_1$  is used to construct an average rate of Type II error. Likewise, a full support prior on the interior of the null set  $\Theta_0$  is used to construct an average rate of Type I error. The test that minimizes the weighted sum (with parameters  $\tau$  and  $1 - \tau$ ) of average Type II and Type I errors in  $\mathcal{C}_{X_2}(\alpha\text{-csb})$  is defined as an ECS test.

Note that ECS tests minimize average risk. For  $\theta \notin \text{Int}\Theta_i$ , set  $p_i(\theta) = 0$ . The function

$$p^*(\theta) \equiv \tau p_1(\theta) + (1 - \tau) p_0(\theta)$$

defines a full-support probability density function on  $\text{Int } \Theta_0 \cup \text{Int } \Theta_1$ . Therefore,  $\phi^*$  is an ECS test if it minimizes average risk, that is,

$$\phi^* \in \arg \min_{\phi \in \mathcal{C}_{X_2}(\alpha\text{-csb})} \int_{\text{Int } \Theta_0 \cup \text{Int } \Theta_1} R(\phi, \theta) p^*(\theta) d\theta.$$

**PRIORS FOR THE GAUSSIAN QUASI-SHIFT MODEL:** Consider again the point-null problem  $\mathbf{H}_0 : \boldsymbol{\mu}_1 = 0$ . The interior of the alternative hypothesis is the alternative

---

<sup>14</sup>A probability density function  $p(\theta)$  is said to have full-support on an open set  $A \subseteq \boldsymbol{\Theta}$  if  $\int_a p(\theta) d\theta > 0$  for all open  $a \subseteq A$ , and  $\int_a p(\theta) d\theta = 0$  for all  $a \in \text{Int}(\boldsymbol{\Theta} \setminus A)$ .

hypothesis itself. The interior of the null hypothesis is empty. Let

$$(z_1, z_2, \dots, z_{n+n^2})' \sim \mathcal{N}_{n+n^2}(\mathbf{0}, \lambda^2 \mathbb{I}_{n+n^2}),$$

where  $\lambda > 0$  is a scalar parameter used to index the priors under study. Consider the following distribution over the parameters of the model:

$$\boldsymbol{\mu}_2 = \begin{bmatrix} z_{n+1} & z_{n+n+1} & \dots & z_{n^2+1} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n+n} & z_{n+n+n} & \dots & z_{n+n^2} \end{bmatrix}$$

and

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2^{-1}(z_1, \dots, z_n)'.$$

Note that  $\boldsymbol{\mu}_2$  has the distribution of a  $n \times n$  random matrix of i.i.d. normal random variables with variance  $\lambda^2$ . Therefore, the inverse  $\boldsymbol{\mu}_2^{-1}$  exists with probability one. The prior over the parameters  $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$  is obtained as a transformation of the multivariate normal vector  $(z_1, \dots, z_{n+n^2})$ .

**Remark 7.** In the point-null problem, the ECS tests are simply maximizers of weighted average power subject to a conditional similarity-on-the-boundary constraint.

### 3.4.2 Theorem 1

Let  $(\mathbf{X}, \Theta, f, \Theta_0)$  be a hypothesis testing problem with a product sample space  $(\mathbf{X}_1, \mathbf{X}_2)$ . Let  $\mathcal{G}$  be a collection of bounded measurable functions,  $g : \mathbf{X} \rightarrow \mathbb{R}$ . Theorem 1 provides a general approach to generating admissible tests within subclasses of the form:

$$\mathcal{C}(\alpha, \mathcal{G}) \equiv \{\phi \in \mathcal{C} \mid \mathbb{E}_\theta[(\phi(x) - \alpha)g(x)] = 0 \quad \forall \theta \in \text{Bd}\Theta_0, g \in \mathcal{G}\}.$$

The suggestion is as follows. First, compute the average rates of Type I/Type II errors with respect to full-support priors. Second, trade off the average rates of Type I/Type



II error using a *strictly monotone function*  $\mathbf{W} : \mathbb{R}^2 \rightarrow \mathbb{R}$ , while imposing the constraints in  $\mathcal{C}(\alpha\text{-}\mathcal{G})$ .<sup>15</sup>

Note that ECS tests are a particular case of this approach. The set of constraints defining the class  $\mathcal{C}_{X_2}(\alpha\text{-csb})$  is given by the collection of all indicator functions of the form:

$$\mathcal{G}_{X_2} = \left\{ g \mid g(x_1, x_2) = 1 \text{ if } x_2 \in \mathcal{F} \text{ and } g(x_1, x_2) = 0 \text{ if } x_2 \notin \mathcal{F}; \mathcal{F} \in \mathcal{B}(X_2) \right\}$$

where  $X_2$  is a boundary-sufficient statistic (see Corollary 2 to Lemma 1 in Appendix A) and  $\mathcal{B}(X_2)$  is the smallest  $\sigma$ -algebra generated by  $X_2$ . In addition, ECS tests use a linear trade-off function  $\mathbf{W}(x, y) = \tau x + (1 - \tau)y$ .

**Theorem 1.** *Let  $p_i(\theta)$  denote a full-support probability density function over  $\text{Int } \Theta_i$ , for  $i = \{0, 1\}$ , and let  $\mathbf{W} : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a continuous, strictly monotone function. Define*

$$M(\mathbf{W}, p_1, p_0, \mathcal{G}) \equiv \arg \min_{\phi \in \mathcal{C}(\alpha\text{-}\mathcal{G})} \mathbf{W} \left( \int_{\text{int } \Theta_1} R(\phi, \theta) p_1(\theta) d\theta, \int_{\text{Int } \Theta_0} R(\phi, \theta) p_0(\theta) d\theta \right).$$

**T1a:** *Suppose that the sample space  $\mathbf{X}$  is topologically separable. Under Assumptions TC1, TC2, and Assumption C,*

$$M(\mathbf{W}, p_1, p_0, \mathcal{G}) \neq \emptyset.$$

**T1b:** *Suppose that  $g^* \in \mathcal{G}$ , where  $g^*(x) = 1$  for all  $x \in \mathbf{X}$ . Under Assumption TC2 and Assumption C,*

$$\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G}) \implies \phi^* \text{ is admissible in } \mathcal{C}(\alpha\text{-}\mathcal{G}).$$

---

<sup>15</sup> $\mathbf{W}(x, y)$  is strictly monotone if whenever  $x \leq x'$ ,  $y \leq y'$  (with either  $x < x'$  or  $y < y'$ ), then  $\mathbf{W}(x, y) < \mathbf{W}(x', y')$ .

**T1c:** Suppose that  $\mathcal{G}^* \equiv \{g^*\}$ . Under Assumption TC1, TC2, and Assumption C,

$$\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G}^*) \implies \phi^* \text{ is admissible in } \mathcal{C}.$$

See Appendix A.3.2 for the proof of Theorem 1.

**Remark 8.** The statement of Theorem 1 is general enough to include testing problems where similarity on the boundary is of interest, but in which there is no boundary-sufficient statistic.<sup>16</sup> The equality  $\mathcal{C}(\alpha\text{-}\mathcal{G}^*) = \mathcal{C}(\alpha\text{-sb})$  implies that such settings are covered by Theorem 1. A very interesting implication of T1a and T1c applied to  $\mathcal{C}(\alpha\text{-}\mathcal{G}^*)$  is the following: *similarity-on-the-boundary is compatible with admissibility* in hypothesis testing problems satisfying Assumptions TC1, TC2, and C. That is to say, there *exists* an admissible test that is similar on the boundary. The test can be obtained as an element of  $M(\mathbf{W}, p_1, p_0, \mathcal{G}^*)$ .<sup>17</sup>

**Remark 9.** Theorem 1 is easily applied to ECS tests by simply noting that  $\mathcal{C}(\alpha\text{-}\mathcal{G}_{X_2})$  is equal to  $\mathcal{C}_{X_2}(\alpha\text{-csb})$ . Theorem 1a guarantees that the ECS tests—which are average risk minimizers—are well-defined. Theorem 1b implies ECS tests are admissible in the class  $\mathcal{C}_{X_2}(\alpha\text{-csb})$ . The admissibility result is extended to all tests using the following condition.

**DEFINITION 5:** (Lehmann and Romano (2005) p. 115) Let  $m : \mathbf{X}_2 \rightarrow \mathbb{R}$  be an arbitrary bounded measurable function. A boundary-sufficient statistic is boundedly complete if

$$\mathbb{E}_{h(\cdot, \theta)}[m(X_2)] = 0, \quad \forall \theta \in \text{Bd}\Theta_0 \implies m(X_2) = 0.$$

---

<sup>16</sup>For example, moment inequality models.

<sup>17</sup>Topological separability of  $\mathbf{X}$  holds automatically in all the applications, since the sample space under consideration is  $\mathbb{R}^s$ .

except, perhaps, in a set that has zero measure under every element of  $\{h(\cdot, \theta)\}_{\theta \in \text{Bd}\Theta_0}$ .

From Theorem 4.3.2 in Lehmann and Romano (2005) it follows that if  $X_2$  is boundedly complete then  $\mathcal{C}(\alpha\text{-}\mathcal{G}^*) = \mathcal{C}(\alpha\text{-}\mathcal{G}_{X_2})$ , or equivalently,  $\mathcal{C}(\alpha\text{-sb}) = \mathcal{C}(\alpha\text{-csb})$ .<sup>18</sup> Then T1c implies that whenever the sufficient statistic  $X_2$  is boundedly complete the ECS tests are admissible within the class of all tests.

**BOUNDED COMPLETENESS IN THE GAUSSIAN QUASI-SHIFT MODEL:** In Section 3.3 it was shown that the gaussian quasi-shift model evaluated at the boundary of null ( $\boldsymbol{\mu}_1 = 0$ ) becomes

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{n+n^2} \left( \begin{pmatrix} \mathbf{0} \\ \text{vec}(\boldsymbol{\mu}_2) \end{pmatrix}, \begin{pmatrix} \mathbb{I}_{n+n^2} \end{pmatrix} \right).$$

Hence, the boundary-sufficient statistic is  $X_2$  and its distribution evaluated at the boundary is given by the statistical model

$$X_2 \sim \mathcal{N}_{n^2}(\text{vec}(\boldsymbol{\mu}_2), \mathbb{I}_{n^2}), \quad \text{vec}(\boldsymbol{\mu}_2) \in \mathbb{R}^{n^2}.$$

Theorem 4.3.1 in Lehmann and Romano (2005) provides a sufficient condition to guarantee that the family of distributions above is complete, and thus, boundedly complete. In this case, it is sufficient to show that the parameter space contains an  $n^2$  dimensional rectangle. That is, the parameter space contains a set of the form:

$$I(a, b) = \{(\mu_{21,1}, \dots, \mu_{21,n}, \mu_{22,1}, \dots, \mu_{22,n}, \dots, \mu_{2n,n})' \in \mathbb{R}^{n^2} \mid a < \mu_{2i,j} < b, \forall i, j\},$$

for  $a, b \in \mathbb{R}, a < b$ .

**Remark 10.** The proof of Theorem 1 uses a novel result concerning the class of  $\alpha$ -

---

<sup>18</sup>See also the Lehmann and Scheffe's Theorem (Linnik (1968) Chapter 4, p. 67).

conditionally (or unconditionally) similar on the boundary tests: *compactness in the weak\* topology* (Lemma 1, Appendix 1). Minimizers of average risk over a compact set  $\mathcal{D}$  (this is, Bayes tests in  $\mathcal{D}$ ) play an important role in *essentially complete class* theorems (see for example, Ferguson (1967), Theorem 2.10.3; Le Cam (1986), Chapter 2, Theorem 1). This is a relevant observation in light of Theorem 1a, which associates different admissible tests with different choices of  $\mathbf{W}$ . An important fact is that ECS tests (plus properly defined limits) form an essentially complete class in the class of  $\alpha$ -conditionally similar-on-the-boundary tests.

### 3.4.3 Theorem 2

Section 3.4.2 presented the main theoretical properties of ECS tests. This section focuses on their implementation. ECS tests were defined as the solution to a minimization problem over a space of functions. Under boundary sufficiency and some regularity conditions, a closed form solution for this problem is available.

INTEGRATED LIKELIHOODS: Let  $p_i(\theta)$  denote a full-support probability density function over  $\text{Int}\Theta_i$ , for  $i = \{0, 1\}$  and let  $\{f(x_1, x_2, \theta)\}_{\theta \in \Theta}$  be a statistical model. Define the *null and alternative integrated likelihoods* as

$$f_i^*(x_1, x_2) \equiv \int_{\text{Int}\Theta_i} f(x_1, x_2, \theta) p_i(\theta) d\theta, \quad i = \{0, 1\}.$$

In cases where  $\text{Int}\Theta_0 = \emptyset$  (i.e.,  $\Theta_0 = \text{Bd}\Theta_0$ ), set  $f_0^*(x_1, x_2) = 0$ .

The key insight of this section is the following. Fubini's Theorem implies that  $\phi^*$  is an ECS test if and only if  $\phi^*$  solves the problem:

$$\min_{\phi \in \mathcal{C}} \quad \tau \int_X (1 - \phi(x)) f_1^*(x) dx + (1 - \tau) \int_X \phi(x) f_0^*(x) dx$$

subject to

$$\int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) f_{\text{Bd}}(x_1|x_2) dx_1 = \alpha.$$

The product structure of  $\mathbf{X}$  and the linearity of the integral can be used to further simplify the objective function:

$$\max_{\phi \in \mathcal{C}} \int_{X_2} \left( \int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) [\tau f_1^*(x_1, x_2) - (1 - \tau) f_0^*(x_1, x_2)] dx_1 \right) dx_2.$$

Hence, it is possible to solve the optimization problem over the functional space by considering the following collection of problems

$$\max_{\phi \in \mathcal{C}} \int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) [\tau f_1^*(x_1, x_2) - (1 - \tau) f_0^*(x_1, x_2)] dx_1$$

subject to

$$\int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) f_{\text{Bd}}(x_1|x_2) dx_1 = \alpha,$$

which can be solved using the Generalized Neyman Pearson Lemma in Ferguson (1967) p. 204. The previous arguments lead to the following definitions.

**ECS TEST STATISTIC:** Let  $\tau \in (0, 1)$ . Define

$$z_{\text{ECS}}(x_1, x_2; p_1, p_0, \tau) \equiv \left[ \tau f_1^*(x_1, x_2) - (1 - \tau) f_0^*(x_1, x_2) \right] / f_{\text{Bd}}(x_1 | x_2).$$

**CRITICAL VALUE FUNCTION:** For each  $x_2 \in \mathbf{X}_2$ , define

$$c(x_2; \alpha) \equiv \arg \min_{q \in \mathbf{X}_1(x_2)} \mathbb{E}_{f_{\text{Bd}}(X_1|x_2)} \left[ \rho_{(1-\alpha)} \left( z_{\text{ECS}}(X_1, x_2; p_1, p_0, \tau) - q \right) \right],$$

where  $\rho_{1-\alpha}(\cdot)$  is the “check function” defined by  $\rho_{1-\alpha}(u) = u[(1 - \alpha) - \mathbb{1}\{u < 0\}]$ . For each  $x_2$ ,  $c(x_2; \alpha)$  corresponds to the conditional  $(1-\alpha)$  quantile of the random variable  $z_{\text{ECS}}(X_1, x_2; p_1, p_0, \tau)$ .

REGULARITY ASSUMPTIONS: The implementation result requires regularity conditions on the integrated and boundary conditional likelihood, but also on the critical value function. A function  $g : X \times Y \rightarrow \mathbb{R}$  is *separately continuous* if  $g(\cdot, y) : X \rightarrow \mathbb{R}$  is continuous for all  $y \in Y$  and  $g(x, \cdot) : Y \rightarrow \mathbb{R}$  is continuous for all  $x \in X$ ; see Rudin (2005) p.52.

**ASSUMPTION R1:**  $f_i^*(x_1, x_2)$ ,  $i = \{0, 1\}$ , and  $f_{\text{Bd}}(x_1|x_2)$  are *separately continuous*.

**ASSUMPTION R2:** The function  $c(\cdot; \alpha) : \mathbf{X}_2 \rightarrow \mathbb{R}$  is measurable.

**Theorem 2.** *Let  $\mathbf{X}$  be separable. Suppose Assumption R1 and R2 hold. Then  $\phi \in M(\tau, p_1, p_0)$  if and only if  $\phi(x)$  equals the test*

$$\phi^*(x_1, x_2) = \mathbb{1}\{z_{\text{ecs}}(x_1, x_2; p_1, p_0, \tau) - c(x_2; \alpha) > 0\},$$

*except, perhaps, in a set of lebesgue-measure zero in  $\mathbf{X}$ .*

See Appendix A.3.3 for the proof of Theorem 2.

Theorem 2 formalizes the arguments presented at the beginning of this section. In principle, the function  $\phi(x_1, x_2)$  that aggregates the “accept-reject” decisions of each conditional optimization problem need not be a well-defined test (i.e., a measurable mapping from  $\mathbf{X}$  to  $[0, 1]$ ). Assumptions R1 and R2 provide a set of sufficient conditions under which the measurability condition is verified.

“POINT” ECS TEST FOR THE GAUSSIAN QUASI-SHIFT MODEL: Consider the testing problem  $\mathbf{H}_0 : \boldsymbol{\mu}_1 = 0$  vs.  $\mathbf{H}_1 : \boldsymbol{\mu}_1 = 0$ . Just as before, let

$$(z_1, z_2, \dots, z_{n+n^2})' \sim \mathcal{N}_{n+n^2}(\mathbf{0}, \lambda^2 \mathbb{I}_{n+n^2}),$$

where  $\lambda > 0$  is a scalar parameter used to index the priors under study. I compute the integrated likelihood  $f_1^*(x_1, x_2)$  under the following priors:

$$\boldsymbol{\mu}_2 = \begin{bmatrix} z_{n+1} & z_{n+n+1} & \cdots & z_{n^2+1} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n+n} & z_{n+n+n} & \cdots & z_{n+n^2} \end{bmatrix}$$

and

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2^{-1}(z_1, \dots, z_n)'.$$

In Section A.3.6 of the Appendix I show that

$$f_1^*(x_1, x_2) = c_1 \exp\left(-\frac{x_1'x_1}{2}\right) \exp\left(\frac{\lambda^2}{2(1+\lambda^2)}x_1'x_1\right) \exp\left(\frac{-x_2'x_2}{2}\right) \exp\left(\frac{\lambda^2}{2(1+\lambda^2)}x_2'x_2\right),$$

where  $c_1$  is a non-negative constant that does not depend on  $(x_1, x_2)$ . Since:

$$f_{\text{Bd}}(x_1|x_2) = c_2 \exp\left(-\frac{1}{2}x_1'x_1\right),$$

the ECS test statistic in Theorem 2 is given by:

$$z(x_1, x_2) \equiv \frac{c_1}{c_2} \exp\left(\frac{1}{2} \frac{\lambda^2}{1+\lambda^2} x_1'x_1\right) \exp\left(-\frac{1}{2}x_2'x_2\right) \exp\left(\frac{1}{2} \frac{\lambda^2}{1+\lambda^2} x_2'x_2\right).$$

Consider the critical value function

$$c(x_2; \alpha) \equiv \frac{c_1}{c_2} \exp\left(\frac{1}{2} \frac{\lambda^2}{1+\lambda^2} \chi_{n,1-\alpha}^2\right) \exp\left(-\frac{1}{2}x_2'x_2\right) \exp\left(\frac{1}{2} \frac{\lambda^2}{1+\lambda^2} x_2'x_2\right),$$

where  $\chi_{n,1-\alpha}^2$  is the  $1 - \alpha$  quantile of a central  $\chi_n^2$  random variable. Note that

$$\mathbb{P}_{f_{\text{Bd}}(x_1|x_2)}(z(x_1, x_2) > c(x_2; \alpha)) = \alpha.$$

Since the function  $c(x_2; \alpha)$  is measurable, as it is continuous in  $x_2$ , Theorem 2 implies that the ECS test rejects if and only if:

$$x_1'x_1 > \chi_{n,1-\alpha}^2,$$

regardless of the parameter  $\lambda^2$ .

**Remark 11.** In the next section I will show that the just-identified IV model with  $n$  endogenous regressors can be reduced to a Gaussian quasi-shift experiment with correlated components; that is

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{n+n^2} \left( \begin{pmatrix} \boldsymbol{\mu}_1\boldsymbol{\mu}_2 \\ \text{vec}(\boldsymbol{\mu}_2) \end{pmatrix}, W \right).$$

The parameter  $\boldsymbol{\mu}_1$  corresponds to the coefficients of the  $n$  (right-hand) endogenous regressors and  $\boldsymbol{\mu}_2$  is the matrix of first-stage coefficients. The random variables  $X_1$  and  $X_2$  are the (standarized and orthogonalized) Ordinary Least Squares estimators of the second-stage and first-stage reduced-form coefficients, respectively. I will show that the  $\alpha$ -ECS test for the just-identified IV model rejects the null hypothesis if  $x_1'x_1 > \chi_{n,1-\alpha}^2$ , which corresponds to the Anderson and Rubin (1949) test.

### 3.5 Examples

This section derives ECS tests in three testing problems that are common in econometric practice: Linear Instrumental Variables regression (IV), a class of weakly identified GMM models with a scalar parameter (wGMM), and Structural Vector Autoregressions (SVARs).

Each example is presented using the following structure:



- (a) Econometric model of interest and description of the testing problem;
- (b) Distributional assumptions;
- (c) Statistical model;
- (d) Boundary sufficiency;
- (e) ECS tests and the priors used to generate them.

### 3.5.1 Linear Instrumental Variables Regression (IV)

This section considers three different set-ups for linear IV regression. First, I study a just-identified IV model with non-stochastic instruments and i.i.d. normal reduced-form errors with a known covariance matrix. I show that the Anderson and Rubin (1949) test (subsequently abbreviated, AR) is ECS (RESULT 1). The priors that generate the AR have an interesting property: the implied distribution for the reduced form parameters have the same law—up to location—as their Ordinary Least-Squares (OLS) estimates.

Second, using the same distributional assumptions as above I study Chamberlain’s (2007) canonical representation of an over-identified IV model with a single endogenous regressor,  $\beta$ . The canonical model has parameters  $(\rho, \phi, \omega)$ , where  $\rho$  is a non-negative scalar measuring the strength of the instruments,  $\phi$  is normalized to be a point on the unit circle, and  $\omega$  is an element on the  $(k - 1)$  sphere that represents the instruments’ direction. I derive a new ECS test for the point hypothesis problem,  $\beta = \beta_0$ , or equivalently  $\phi_1 = 0$  (RESULT 2). The priors for  $\phi$  and  $\omega$  are uniform on their domain and  $\rho \sim \sqrt{\lambda^2 \chi_k^2}$ , where  $\lambda^2$  is a free parameter for the researcher.<sup>19</sup> The ECS test for these priors depends on the maximal invariant in Andrews *et al.* (2006). The test has

---

<sup>19</sup>Chamberlain (2007) shows that the choice of prior distributions for  $\phi$  and  $\omega$  arise from a solution to a minimax problem.

optimality properties that neither Moreira’s (2003) CLR, nor the Wald tests based on TSLS or LIML have been shown to satisfy; namely admissibility in the class of all tests and efficiency in the class of similar tests.

Third, I study a just-identified IV model using the “local-to-zero” framework of Staiger and Stock (1997). A weak convergence result for the reduced-form OLS coefficients provides a limiting experiment—in the sense of Müller (2011)—that is convenient for the study of just-identified IV models with heteroskedastic, autocorrelated, and/or clustered data, all of which are common features in applied work. Once-again, a “robust” version of the AR test—which incorporates the asymptotic variance of the OLS coefficients—is shown to be ECS.

## General Gaussian IV model

**a) ECONOMETRIC MODEL:** Consider a linear IV model in reduced form matrix notation with  $n$  endogenous regressors and  $k \geq n$  instruments. The notation follows the simultaneous equations framework of Moreira (2003),

$$\begin{aligned} y_1 &= Z\Pi\beta + v_1, \\ Y_2 &= Z\Pi + V_2. \end{aligned}$$

The structural parameter of interest is  $\beta \in \mathbb{R}^n$ , while  $\Pi \in \mathbb{R}^{k \times n}$  denotes the unknown matrix of first-stage coefficients. The sample size is  $T$  and the econometrician observes the data set  $\{y_{1t}, Y_{2t}, Z_t\}_{t=1}^T$ . I denote observations of the outcome variable, the  $n$  endogenous regressors, and the vector of instruments by  $y_{1t}$ ,  $Y_{2t}$  and  $Z_t$ , respectively. The unobserved reduced form errors have realizations  $v_{1t}$  and  $V_{2t}$ . I stack the realized

variables in matrices  $y_1 \in \mathbb{R}^T$ ,  $Y_2 \in \mathbb{R}^{T \times n}$ ,  $Z \in \mathbb{R}^{T \times k}$ ,  $v_1 \in \mathbb{R}^T$ ,  $V_2 \in \mathbb{R}^{T \times n}$ . The testing problem of interest is

$$\mathbf{H}_0 : \beta = \beta_0 \quad vs. \quad \mathbf{H}_1 : \beta \neq \beta_0.$$

**b) DISTRIBUTIONAL ASSUMPTIONS:** Assume that the  $T$  rows of the  $T \times (n+1)$  matrix of reduced-form errors  $V = [v_1, V_2]$  are i.i.d. normally distributed with mean zero and known nonsingular covariance matrix  $\Omega = [\omega_{ij}]$ . This is,

$$\text{vec}(V) \sim \mathcal{N}_{T(n+1)}(\mathbf{0}, \Omega \otimes \mathbb{I}_T),$$

where “ $\otimes$ ” denotes the Kronecker product. For simplicity, assume that  $Z$  is non-stochastic.

**c) STATISTICAL MODEL:** Let  $Y = [y_1, Y_2]$ . Under the normality assumption for  $V$ , the sufficient statistics for the IV model are the reduced-form ordinary least-squares (OLS) estimates of  $\Pi\beta$  and  $\Pi$ ,

$$\hat{\gamma}_{\text{OLS}} \equiv \text{vec}((Z'Z)^{-1}Z'Y) \sim \mathcal{N}_{k+nk} \left( \begin{pmatrix} \Pi\beta \\ \text{vec}(\Pi) \end{pmatrix}, \Omega \otimes (Z'Z)^{-1} \right).$$

**d) BOUNDARY SUFFICIENCY:** The boundary of the null hypothesis  $\mathbf{H}_0 : \beta = \beta_0$  is the null hypothesis itself

$$\text{Bd}\Theta_0 = \{(\beta, \Pi) \in \mathbb{R}^n \times \mathbb{R}^{k \times n} \mid \beta = \beta_0\}.$$

In Appendix A.3.4, I show that the following rotation and standarization of the OLS coefficients:

$$\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \equiv (C_0 \otimes (Z'Z)^{1/2}) \hat{\gamma}_{\text{OLS}}$$

yields a statistical model in which  $\gamma_2^*$  is a boundary-sufficient statistic. The transformation follows Moreira (2003), p. 1030 with

$$C_0 \equiv \begin{pmatrix} (b_0' \Omega b_0)^{-1/2} b_0' \\ (A_0' \Omega^{-1} A_0)^{-1/2} A_0' \Omega^{-1} \end{pmatrix},$$

$$b_0 = [1, -\beta_0']' \quad A_0 = [\beta_0, \mathbb{I}_n]'$$

Intuitively,  $\gamma_2^*$  corresponds to the standardized and normalized coefficients from the first-stage regressions.

### Just-identified Gaussian Model

**e1)** PRIORS FOR THE JUST-IDENTIFIED MODEL AND ECS TEST: An IV model is just-identified if  $k = n$ . Consider the following multivariate normal vector

$$\gamma \sim \mathcal{N}_{n+n^2}(\mathbf{0}, \Omega \otimes (Z'Z)^{-1}).$$

Write  $\gamma = (\gamma_1', \gamma_{21}', \gamma_{22}', \dots, \gamma_{2n}')'$  where  $\gamma_1$  is  $n \times 1$  and  $\gamma_{2i}$  is  $n \times 1$  for all  $i = 1, \dots, n$ . Consider the following prior distribution over the parameters  $(\beta, \Pi)$ , which are natural extensions of the priors used in the Gaussian quasi-shift model in the previous section:

$$\Pi = [\gamma_{21}, \gamma_{22}, \dots, \gamma_{2n}],$$

$$\beta = [\gamma_{21}, \gamma_{22}, \dots, \gamma_{2n}]^{-1} \gamma_1.$$

Note that  $\Pi$  has the distribution of a Gaussian random matrix, and that  $\Pi$  is full rank with probability one, provided the covariance matrix  $\Omega \otimes (Z'Z)^{-1}$  is nonsingular. In light of this observation, the distribution for  $\beta$  is well-defined. There is an interesting feature about the distribution selected: it is the unique distribution for which the reduced form coefficients  $(\Pi\beta, \Pi)$  have the same random behavior—up to location—as their sample counterparts,

$$\begin{pmatrix} \Pi\beta \\ \text{vec}(\Pi) \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \text{vec}(\gamma_{21}, \gamma_{22}, \dots, \gamma_{2n}) \end{pmatrix} \sim \mathcal{N}_{n+n^2}(\mathbf{0}, \Omega \otimes (Z'Z)^{-1}).$$

**Result 1.** *The  $\alpha$ -ECS test for the problem  $\mathbf{H}_0 : \beta = \beta_0$  vs.  $\mathbf{H}_1 : \beta \neq \beta_0$  in a just-identified IV model with  $n$  endogenous regressors and the priors in e1) rejects the null if*

$$\gamma_1^{*'} \gamma_1^* = (y_1 - Y_2\beta_0)' Z(Z'Z)^{-1} Z'(y_1 - Y_2\beta_0) / (b_0' \Omega b_0) > \chi_{n,1-\alpha}^2.$$

*Hence, the Anderson and Rubin (1949) test is efficient conditionally similar on the boundary.*

See Appendix A.3.5 for the proof of Result 1.

**Remark 12.** Note that  $\gamma_2^*$  is boundedly complete. This follows from the fact that the collection of distributions

$$\mathcal{N}_{n^2} \left( \text{vec}[(Z'Z)^{1/2} \Pi (A_0' \Omega^{-1} A_0)^{1/2}] , \mathbb{I}_{n^2} \right), \quad \Pi \in \mathbb{R}^{n \times n},$$

is complete. Consequently, Result 1 provides a new sense of optimality for the AR test; namely, *efficiency*: the test maximizes weighted average power—with respect to the full-support priors described above—among all similar tests. This observation complements previous results in the literature. For instance, Theorem 3 in Moreira (2009) shows that the AR test is *uniformly most powerful* among the class of *unbiased tests* (UMPB), provided the IV model is just-identified and there is a single endogenous

regressor. Result 1 applies to a larger class of tests (similar tests) and to a larger class of IV models (just-identified, arbitrary number of endogenous regressors).

**Remark 13.** Remark 6, Result 1, and Theorem 1c imply that the AR test is admissible within the class of *all* tests in the context of a Gaussian just-identified model with an arbitrary number of endogenous regressors. Therefore, Result 1 complements Corollary 2 to Theorem 1 in Chernozhukov *et al.* (2009), which shows that the AR test is  $\alpha$ -admissible in Gaussian over-identified models with a single endogenous regressor.<sup>20</sup>

### Over-Identified Gaussian Linear IV model

Chamberlain (2007) introduces a *canonical representation* of the Gaussian IV model with a single endogenous regressor ( $n = 1$ ):

$$\begin{pmatrix} S \\ T \end{pmatrix} \equiv \begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \sim \mathcal{N}_{2k} \left( \begin{matrix} \rho(\phi \otimes \omega), & \mathbb{I}_{2k} \end{matrix} \right).$$

The sample space is  $\mathbb{R}^{2k}$  and the parameter space is as follows

$$\rho \in \mathbb{R}_+, \quad \phi \in \mathcal{S}^1(r(\beta_0)), \quad \omega \in \mathcal{S}^{k-1},$$

where  $\mathcal{S}^m$  is the  $m$  unit sphere; that is,  $\mathcal{S}^m = \{x \in \mathbb{R}^{m+1} \mid \|x\| = 1\}$ , for any  $m \in \mathbb{N}$  and

$$\mathcal{S}^1(r(\beta_0)) = \{(\phi_1, \phi_2) \in \mathcal{S}^1 \mid r(\beta_0)\phi_1 + \sqrt{1 - r^2(\beta_0)}\phi_2 \geq 0\},$$

with  $r(\beta_0)$  equal to the correlation coefficient of the  $2 \times 2$  matrix  $b'_0 \Omega b_0$ .

---

<sup>20</sup>See Lehmann and Romano (2005), p. 233 for the differences between  $\alpha$ -admissibility and admissibility.

The original parameters  $(\beta, \Pi)$  induce the following canonical parameters  $(\rho, \phi, \omega)$ :

$$\rho = (A'\Omega^{-1}A)^{1/2}(\Pi'Z'Z\Pi)^{1/2}, \quad \phi = C_0A/(A'\Omega^{-1}A)^{1/2}, \quad \omega = (Z'Z)^{1/2}\Pi/(\Pi'Z'Z\Pi)^{1/2},$$

where  $A \equiv [\beta, 1]'$ , and  $C_0$  is the  $2 \times 2$  matrix with first row equal to  $(b'_0\Omega b_0)^{-1/2}b'_0$  and second row given by  $(A'_0\Omega^{-1}A_0)^{-1/2}A'_0\Omega^{-1}$  defined in the previous section.

**d) BOUNDARY SUFFICIENCY IN THE CANONICAL MODEL:** The testing problem

$$\mathbf{H}_0 : \beta \leq (=) \beta_0 \quad vs. \quad \mathbf{H}_1 : \beta \not\leq (\neq) \beta_0,$$

is equivalent to

$$\mathbf{H}_0 : \phi_1 \leq (=) 0 \quad vs. \quad \mathbf{H}_1 : \phi_1 \not\leq (\neq) 0.$$

Therefore, on the boundary of the null hypothesis

$$\text{Bd}\Theta_0 = \{(\rho, \phi, \omega) \in \mathbb{R}_+ \times \mathcal{S}^1(r(\beta_0)) \times \mathcal{S}^{k-1} \mid \phi_1 = 0\},$$

the canonical statistical model becomes:

$$\begin{pmatrix} S \\ T \end{pmatrix} \sim \mathcal{N}_{2k} \left( \begin{pmatrix} 0 \\ \rho\omega \end{pmatrix}, \mathbb{I}_{2k} \right).$$

Hence,  $T$  is a boundary-sufficient statistic. Note that  $T$  is also boundedly complete.

**e2) PRIORS FOR THE OVER-IDENTIFIED IV MODEL ( $\mathbf{H}_0 : \phi_1 = 0$ ):** Following Chamberlain (2007) the parameters  $(\rho, \phi, \omega)$  are treated as independent random variables. The distributions for  $\phi$  and  $\omega$  are uniform on their domain:

$$\phi \sim \mathcal{U}(\mathcal{S}^1(r(\beta_0))), \quad \omega \sim \mathcal{U}(\mathcal{S}^{k-1}).$$

Chamberlain (2007) shows that these choices arise from the solution of a minimax

problem. I consider the following prior over the parameter  $\rho$  :

$$\rho \sim \sqrt{\lambda^2 \chi_k^2}.$$

**Result 2.** *The  $\alpha$ -ECS test for the problem  $\mathbf{H}_0 : \phi_1 = 0$  vs.  $\mathbf{H}_1 : \phi_1 \neq 0$  in an over-identified IV model with a single endogenous regressor and the priors in e2) rejects the null hypothesis if the statistic:*

$$(S'S - T'T) + \frac{4(1 + \lambda^2)}{\lambda^2} \ln \left[ I_0 \left( \frac{\lambda^2}{4(1 + \lambda^2)} \left[ (S'S - T'T)^2 + 4(S'T)^2 \right]^{1/2} \right) \right]$$

*exceeds the critical value function  $c^*(T; \lambda^2, \alpha)$ , defined as the  $(1 - \alpha)$  quantile of the distribution of the statistic above with  $S \sim \mathcal{N}_k(\mathbf{0}, \mathbb{I}_k)$  and  $T$  fixed. The function  $I_0(\cdot)$  is the modified Bessel function of the first kind of order zero defined in Section 9.6, p. 375 of Abramowitz and Stegun (1964).*

See Appendix A.3.6 for a proof of Result 2.

**Remark 14.** Let  $AR \equiv S'S$  denote the Anderson and Rubin (1949) statistic for the over-identified IV model.<sup>21</sup> Let  $LM \equiv (S'T)^2/T'T$  denote the Lagrange Multiplier statistic as defined in Andrews *et al.* (2006), p. 722. The ECS test in Result 4 is measurable with respect to the triplet  $(AR, LM, T'T)$ . Hence, it is natural to ask whether the ECS test rejects the null hypothesis when both the AR and LM do.

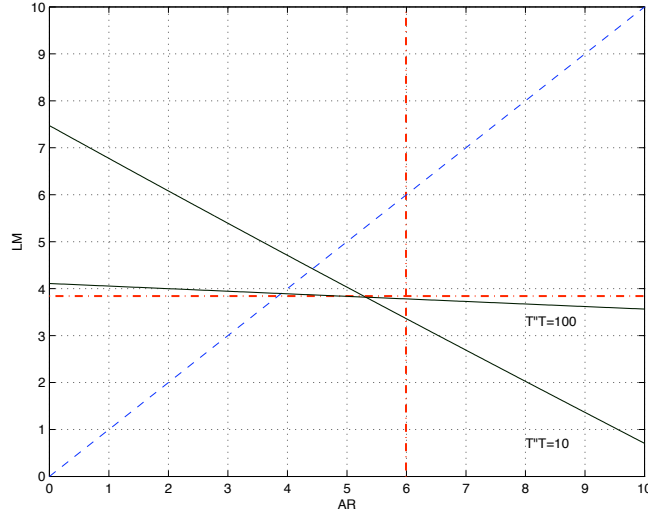
Figure ?? reports “conditional” critical regions in the  $(AR, LM)$  space for two different values of  $T'T$ . The conditional critical region is the collection of  $(AR, LM)$  points at the right of the black (solid) lines (large AR and large LM). Each solid line traces the boundary of the rejection region of the ECS test for a given value of  $T'T \in \{10, 100\}$ .<sup>22</sup>

---

<sup>21</sup>This is a slight abuse of notation as  $AR = S'S/k$ ; see Andrews *et al.* (2006).

<sup>22</sup>The command `ezplot` in Matlab is used to graph the solution to the equation  $z(AR, LM, T'T; \lambda^2) - c(T'T; \lambda^2) = 0$ .





(BLUE, DASHED) Boundary of the sample space:  $AR \geq LM$ . (RED, DOT-DASHED) 5% critical values for the AR and the LM statistics obtained as the upper 5% quantiles of the distributions  $\chi^2_2$  and  $\chi^2_1$ , respectively.

**Figure 3.1:** 5% Conditional Critical Region  $(AR, LM)$ ;  $k = 2$ ,  $\lambda^2 = 1$

The black solid line close to the LM critical value corresponds to the highest realization of  $T'T$ . The ECS test adjusts the  $\chi^2_k$  threshold for the AR depending on the realizations of LM. Interestingly, the magnitude of the adjustment depends on the observed value of the boundary-sufficient statistic. For example, suppose LM is close to one and  $T'T = 10$ . The  $\chi^2_{k,5\%}$  critical value for the AR is adjusted upwards and the null hypothesis is rejected only if  $AR > 9.7 > \chi^2_{k,5\%}$ . If, however,  $T'T = 100$  the adjustment required is significantly larger. The “conditional” critical regions depicted in Figure 1 suggest that the ECS test in Result 2 rejects the null hypothesis whenever  $LM > \chi^2_{1,5\%}$ , provided  $T'T$  is large.

Result 2\* in Section A.3.7 of the Appendix presents a new test for the hypothesis  $\beta \leq \beta_0$ , or equivalently, for  $\phi_1 \leq 0$  in the canonical model.

## (Weakly) Just-Identified IV

This section extends the results in Section 3.5.1 to weakly just-identified IV models outside the conditionally homoskedastic, serially uncorrelated framework. This generalization is relevant for practitioners working with heteroskedastic, time series, or panel data. The arguments in this section are analogous to those made in the just-identified Gaussian model. However, it is important to keep the examples in two different sections. The objective is to illustrate the difference between a statistical model generated by a set of finite-sample distributional assumptions and one generated by a set of (asymptotic) weak convergence assumptions.

**b) DISTRIBUTIONAL ASSUMPTIONS:** I consider the following set of weak convergence assumptions

(1) *Weak Convergence:*  $\text{vec}(\sqrt{T}(Z'Z)^{-1}Z'V) \xrightarrow{d} \mathcal{N}_{n+n^2}(\mathbf{0}, W)$ , where  $W$  is a known nonsingular covariance matrix of dimension  $(n + n^2) \times (n + n^2)$ .

(2) *Local to zero:*  $\Pi = C/\sqrt{T}$ , where  $C$  is a  $n \times n$  matrix.

**c) STATISTICAL MODEL:** The set of weak convergence assumptions induce the following limiting experiment [in the sense of Müller (2011)],

$$\sqrt{T}\hat{\gamma}_{\text{OLS}} \equiv \text{vec}(\sqrt{T}(Z'Z)^{-1}Z'Y) \xrightarrow{d} \mathcal{N}_{n+n^2} \left( \begin{pmatrix} C\beta \\ \text{vec}(C) \end{pmatrix}, W \right).$$

Therefore, the main difference in the limiting statistical model is that the covariance matrix  $W$  need not have a kronecker structure.<sup>23</sup> Rotate the model by the matrix:

---

<sup>23</sup>Later, I will argue that this is an important statistical feature. For instance, the model loses invariance to certain groups of orthonormal rotations.

$$D_0 \equiv \begin{pmatrix} b'_0 \\ A'_0 \end{pmatrix} \otimes \mathbb{I}_n$$

so that

$$D_0 \sqrt{T} \hat{\gamma}_{\text{OLS}} \xrightarrow{d} \begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \sim \mathcal{N}_{n+n^2} \left( \begin{pmatrix} C(\beta - \beta_0) \\ \text{vec}(CA'A_0) \end{pmatrix}, D_0 W D'_0 \right),$$

where

$$D_0 \sqrt{T} \hat{\gamma}_{\text{OLS}} = \begin{pmatrix} \sqrt{T}(\hat{\gamma}_1 - \hat{\gamma}_2 \beta_0) \\ \text{vec}(\sqrt{T}[\hat{\gamma}_1, \hat{\gamma}_2]A_0) \end{pmatrix},$$

and  $\hat{\gamma}_1, \hat{\gamma}_2$  are the OLS estimates of  $\Pi\beta$  and  $\Pi$  respectively.

**d) BOUNDARY SUFFICIENCY:** The limiting experiment has two parameters: the structural parameter of interest,  $\beta$ , and the “drift” parameter  $C$ . The sample space is  $\mathbb{R}^{n+n^2}$ . In order to establish boundary sufficiency it is necessary to further transform the model. Let  $[D_0 W D'_0]_n^{-1/2}$  represent the square root of the inverse of the first  $n \times n$  block of the matrix  $D_0 W D'_0$ . In section A.3.8 of the Appendix, I show that there is a  $n + n^2$  square matrix of the form

$$D \equiv \begin{pmatrix} [D_0 W D'_0]_n^{-1/2} & \mathbf{0} \\ d_1 & d_2 \end{pmatrix},$$

such that  $D(D_0 W D'_0)D' = \mathbb{I}_{n+n^2}$ , where  $d_1$  is a  $n^2 \times n$  matrix and  $d_2$  is  $n^2 \times n^2$ . Note that

$$D \begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} = \begin{pmatrix} [D_0 W D'_0]_n^{-1/2} \gamma_1^* \\ d_1 \gamma_1^* + d_2 \gamma_2^* \end{pmatrix} \sim \mathcal{N}_{n+n^2} \left( D \begin{pmatrix} C(\beta - \beta_0) \\ \text{vec}(CA'A_0) \end{pmatrix}, \mathbb{I}_{n+n^2} \right).$$

Therefore, whenever  $\beta = \beta_0$ , any change in the drift parameter  $C$  affects the distribu-

tion of  $D(\gamma_1^{*'}, \gamma_2^{*'})'$  only through its effect on  $d_1\gamma_1^* + d_2\gamma_2^*$ , which is the boundary-sufficient statistic for the model.

**e3)** PRIORS FOR THE (WEAKLY) JUST-IDENTIFIED MODEL: There is a natural extension for the priors postulated for the Gaussian model in the previous section. Consider the following multivariate normal vector

$$\gamma \sim \mathcal{N}_{n+n^2}(\mathbf{0}, W).$$

Write  $\gamma = (\gamma_1', \gamma_{21}', \gamma_{22}', \dots, \gamma_{2n}')'$  where  $\gamma_1$  is  $n \times 1$  and  $\gamma_{2i}$  is  $n \times 1$  for all  $i = 1, \dots, n$ , and consider the following prior distribution over the parameters  $(\beta, C)$

$$C = [\gamma_{21}, \gamma_{22}, \dots, \gamma_{2n}],$$

$$\beta = [\gamma_{21}, \gamma_{22}, \dots, \gamma_{2n}]^{-1} \gamma_1.$$

RESULT 1\* *The  $\alpha$ -ECS test for the problem  $\mathbf{H}_0 : \beta = \beta_0$  vs.  $\mathbf{H}_1 : \beta \neq \beta_0$  in the limiting experiment of a weakly just-identified IV model with  $n$  endogenous regressors and the priors in e3) rejects the null hypothesis if*

$$\gamma_1^{*'} [D_0 W D_0']_n^{-1} \gamma_1^* > \chi_{n,1-\alpha}^2.$$

Hence, the test evaluated at sample analogues rejects if

$$T(\hat{\gamma}_1 - \hat{\gamma}_2 \beta_0)' [D_0 W D_0']_n^{-1} (\hat{\gamma}_1 - \hat{\gamma}_2 \beta_0) > \chi_{n,(1-\alpha)}^2$$

where  $\hat{\gamma}_1$  is the OLS estimate of  $\Pi\beta$  and  $\hat{\gamma}_2$  is the OLS estimate of the matrix  $\Pi$ . The matrix  $W$  is the asymptotic variance of  $\sqrt{T}\text{vec}[\hat{\gamma}_1, \hat{\gamma}_2]$ , the matrix  $D_0$  is defined in **c)**, and  $[D_0 W D_0']_n$  is the first  $n \times n$  block of the matrix  $D_0 W D_0'$ . The test in Result 1 is simply a robust version of the AR test as

$$\hat{\gamma}_1 - \hat{\gamma}_2 \beta_0 = (Z'Z)^{-1} Z'(y - Y\beta_0).$$

### 3.5.2 Weakly Identified GMM Models

In this section, I shall derive point ECS tests for weakly identified GMM models. The limiting experiment for this problem is based on the following observation: both the sample moment condition of a weakly identified GMM model and its derivative are asymptotically normal in large samples, provided both objects are evaluated at the boundary of the null hypothesis. The location parameter of the limiting normal distribution depends on the shape of the population moment function. Hence the limiting experiment of a weakly identified GMM model exhibits, in principle, an infinite dimensional nuisance parameter. I study problems in which the population moment function is known up to a finite-dimensional vector (as, for example, in an IV model with heteroskedastic and/or serially correlated errors).

There are two results in this section. First, I provide sufficient conditions under which the  $S$ -test of Stock and Wright (2000) is ECS. Second, I provide a general expression for ECS tests in a more general class of models. The concepts and main results in this section are illustrated using a weakly identified IV model with non-homoskedastic and/or serially correlated errors

**a) ECONOMETRIC MODEL:** Let  $x_t$  be an  $\mathbb{R}^d$ -valued random variable. The econometrician observes the data set  $\{x_t\}_{t=1}^T$ , whose unknown distribution depends on a scalar parameter  $\theta \in \mathbb{R}$ . There is a known  $\mathbb{R}^m$ -valued function  $h(x_t, \theta)$  that identifies the true parameter  $\theta^*$  through the following moment condition:

$$\mathbb{E}_{\theta^*}[h(x_t, \theta)] = \mathbf{0} \quad \text{only at } \theta = \theta^*, \forall t \quad (\text{Global Identification}).$$

I assume the function  $h(x_t, \theta)$  is almost-surely differentiable with respect to  $\theta$ , with derivative  $\dot{h}(x_t, \cdot) \equiv \partial h(x_t, \theta) / \partial \theta$  and that

$$\partial \mathbb{E}_{\theta^*}[h(x_t, \theta)] / \partial \theta = \mathbb{E}_{\theta^*}[\dot{h}(x_t, \theta)].$$

The testing problem of interest is

$$\mathbf{H}_0 : \theta^* = \theta_0 \quad vs. \quad \mathbf{H}_1 : \theta^* \neq \theta_0.$$

EXAMPLE (GMM-IV): Let  $x_t \equiv (y_t, Y_t, Z_t)$ , where  $y_t$  is the outcome variable;  $Y_t$  is a single endogenous regressor, and  $Z_t$  is a vector of  $k \times 1$  instruments. Consider the function

$$h(y_t, Y_t, Z_t, \theta) = Z_t(y_t - \theta Y_t).$$

Note that

$$\mathbb{E}_{\theta^*}[h(y_t, Y_t, Z_t, \theta_0)] = (\theta^* - \theta_0)\mathbb{E}[Z_t Y_t] = (\theta^* - \theta_0)\mathbb{E}[Z_t Z_t']\Pi$$

and

$$\mathbb{E}_{\theta^*}[\dot{h}(x_t, \theta)] = -\mathbb{E}[Z_t Z_t']\Pi.$$

**b) DISTRIBUTIONAL ASSUMPTIONS:** Stock and Wright (2000) developed nonstandard asymptotic theory for models defined by moment conditions when some or all of the parameters are weakly identified. I shall use their asymptotic framework to derive a limiting experiment as defined by Müller (2011). Consider the following set of (point-wise in  $\theta^*$ ) weak convergence assumptions for the sample moment condition and its derivative:

$$\begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^T \left( h(x_t, \theta_0) - \mathbb{E}_{\theta^*}[h(x_t, \theta_0)] \right) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T \left( \dot{h}(x_t, \theta_0) - \mathbb{E}_{\theta^*}[\dot{h}(x_t, \theta_0)] \right) \end{bmatrix} \xrightarrow{d} \mathcal{N}_{2m}(\mathbf{0}, \Omega(\theta_0)) \quad \forall \theta^*.$$

To model weak-identification, assume

$$\mathbb{E}_{\theta^*}[h(x_t, \theta_0)] = C_t(\theta^*, \theta_0, \delta)/\sqrt{T},$$

where  $C_t$  is known up to the finite-dimensional nuisance parameter  $\delta \in \mathbb{R}^n$ . The global identification assumption implies  $C_t(\theta_0, \theta_0, \delta) = 0$  for all  $t$ , regardless of the value of the nuisance parameter  $\delta$ . Consider the following regularity conditions for  $C_t$  and its derivative  $\dot{C}_t(\theta^*, \theta, \delta) \equiv \partial C_t(\theta^*, \theta, \delta)/\partial \theta$ :

$$C(\theta^*, \theta_0, \delta) \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T C_t(\theta^*, \theta_0, \delta) < \infty,$$

$$\dot{C}(\theta^*, \theta_0, \delta) \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \dot{C}_t(\theta^*, \theta_0, \delta) < \infty.$$

EXAMPLE (GMM-IV): Using the local-to-zero assumption of Staiger and Stock (1997),

$$\mathbb{E}_{\theta^*}[h(y_t, Y_t, Z_t, \theta_0)] = (\theta^* - \theta_0)\mathbb{E}[Z_t Z_t']\Pi = (\theta^* - \theta_0)\mathbb{E}[Z_t Z_t']\delta/\sqrt{T}$$

and

$$\mathbb{E}_{\theta^*}[\dot{h}(y_t, Y_t, Z_t, \theta_0)] = -\mathbb{E}[Z_t Z_t']\Pi = -\mathbb{E}[Z_t Z_t']\delta/\sqrt{T}.$$

Therefore,

$$C_t(\theta^*, \theta_0, \delta) = (\theta^* - \theta_0)\mathbb{E}[Z_t Z_t']\delta$$

and

$$\dot{C}_t(\theta^*, \theta_0, \delta) = -\mathbb{E}[Z_t Z_t']\delta.$$

Under standard regularity conditions for the second moments of  $Z_t$ ,

$$C(\theta^*, \theta_0, \delta) = (\theta^* - \theta_0) \delta \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E[Z_t Z_t'] < \infty,$$

$$\dot{C}(\theta^*, \theta_0, \delta) = -\delta \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E[Z_t Z_t'] < \infty.$$

c) STATISTICAL MODEL: The set of weak convergence assumptions and the weak identification condition yield the following limiting statistical model for the GMM problem:

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^T h(x_t, \theta_0) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T \dot{h}(x_t, \theta_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}_{2m} \left( \begin{pmatrix} C(\theta^*, \theta_0, \delta) \\ C'(\theta^*, \theta_0, \delta) \end{pmatrix}, \Omega(\theta_0) \right) \quad \forall \theta^*.$$

EXAMPLE (GMM-IV): The limiting experiment for the GMM-IV model with a single endogenous regressor is given by

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_t (y_t - \theta_0 Y_t) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_t Y_t \end{pmatrix} \xrightarrow{d} \mathcal{N}_{2k} \left( \begin{pmatrix} (\theta^* - \theta_0) Q \delta \\ Q \delta \end{pmatrix}, \Omega(\theta_0) \right),$$

where

$$Q \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E[Z_t Z_t'].$$

Since  $Q$  is assumed known and nonsingular in the limiting experiment it is possible to redefine  $\tilde{\delta}$  as  $Q\delta$ . In a slight abuse of notation  $\tilde{\delta}$  is relabeled as  $\delta$ . The specific form of the matrix  $\Omega(\theta_0)$  depends on primitive assumptions about the data. Suppose for simplicity that  $\{y_t, Y_t, Z_t\}$  is obtained from an independent sample with heteroskedasticity. In



that case

$$\Omega(\theta_0) = \begin{pmatrix} \lim_{t \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[ Z_t Z_t' (y_t - \theta_0 Y_t)^2 \right] & \lim_{t \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[ Z_t Z_t' (y_t - \theta_0 Y_t) v_{2t} \right] \\ \lim_{t \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[ Z_t Z_t' (y_t - \theta_0 Y_t) v_{2t} \right] & \lim_{t \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[ Z_t Z_t' v_{2t}^2 \right] \end{pmatrix}.$$

Therefore, the limiting distribution of the sample moment condition for a linear IV model is Gaussian centered at  $(\theta^* - \theta_0)\delta$ . The limiting distribution of the derivative of the sample moment function is also Gaussian, but centered at  $-\delta$ . The components are jointly normal and their dependence structure changes depending on whether the data is heteroskedastic, autocorrelated, or clustered.

**d) BOUNDARY SUFFICIENCY:** In order to establish the existence of a boundary-sufficient statistic, I will rotate and standarize the limiting experiment described above. Let  $[\Omega(\theta_0)]_m$  denote the upper left  $m \times m$  block of the matrix  $\Omega(\theta_0)$ ; that is, the asymptotic variance of the sample moment condition. In section A.3.8 of the Appendix, I show there is a  $2m$  square matrix of the form:

$$D(\theta_0) \equiv \begin{pmatrix} [\Omega(\theta_0)]_m^{-1/2} & \mathbf{0} \\ d_1 & d_2 \end{pmatrix},$$

such that  $D(\theta_0)\Omega(\theta_0)D(\theta_0)' = \mathbb{I}_{2m}$ , where  $d_1$  and  $d_2$  are  $m \times m$  matrices. Therefore

$$\begin{aligned}
\begin{pmatrix} m_T(\theta_0) \\ d_T(\theta_0) \end{pmatrix} &\equiv D(\theta_0) \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^T h(x_t, \theta_0) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T \dot{h}(x_t, \theta_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} m(\theta_0) \\ d(\theta_0) \end{pmatrix} \\
&\sim \mathcal{N}_{2m} \left( \begin{pmatrix} [\Omega(\theta_0)]_m^{-1/2} C(\theta^*, \theta_0, \delta) \\ d_1 C(\theta^*, \theta_0, \delta) + d_2 \dot{C}(\theta^*, \theta_0, \delta) \end{pmatrix}, \mathbb{I}_{2m} \right).
\end{aligned}$$

Thus, the limiting experiment of a weakly identified GMM model has the following features. The sample space is  $\mathbb{R}^{2m}$ : the set of possible values for the vector  $(m(\theta_0)', d(\theta_0)')$ . The parameter space is  $\mathbb{R}^{n+1}$ : the set of possible values for the parameter of interest  $\theta^*$  and the nuisance vector  $\delta$ . The statistical model is a Gaussian Location problem with independent components.

The global identification assumption implies that whenever  $\theta^* = \theta_0$

$$\begin{pmatrix} m(\theta_0) \\ d(\theta_0) \end{pmatrix} \sim \mathcal{N}_{2m} \left( \begin{pmatrix} \mathbf{0} \\ d_2 \dot{C}(\theta^*, \theta_0, \delta) \end{pmatrix}, \mathbb{I}_{2m} \right).$$

Thus,  $d(\theta_0)$  is a boundary sufficient statistic in the limiting experiment of the weakly identified GMM model.

**e) PRIORS FOR THE WEAKLY IDENTIFIED GMM MODEL:** First, I will provide sufficient conditions under which the  $S$ -test of Stock and Wright (2000)—based on the

continuously updated GMM objective function—is ECS. For a fixed  $\theta_0$ , consider the mapping  $C^* : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{2m}$  given by

$$C^*(\theta^*, \delta) = \begin{pmatrix} C(\theta^*, \theta_0, \delta) \\ \dot{C}(\theta^*, \theta_0, \delta) \end{pmatrix}.$$

**ASSUMPTION RGMM1:**  $C^*$  is a continuous function.

EXAMPLE (GMM-IV): In the GMM-IV model, the dimension of the nuisance parameter ( $n$ ) equals the number of instruments ( $k$ ). Likewise, the dimension of the moment conditions ( $m$ ) equals  $k$ . The mapping  $C^* : \mathbb{R}^{k+1} \rightarrow \mathbb{R}^{2k}$  is given by

$$\begin{pmatrix} (\theta^* - \theta_0)\delta \\ \delta \end{pmatrix}.$$

The mapping  $C^*$  is continuous in  $(\theta^*, \delta)$ . Note that  $C^*$  is crucial for comparing the power performance of different testing procedures. The following result states sufficient conditions under which the  $S$ -test of Stock and Wright (2000) is ECS.

**Result 3.** *Let  $n + 1 \geq 2m$  and let assumption RGMM1 hold. Suppose that there is a full-support prior  $p_1$  over  $\mathbb{R}^{n+1}$  such that:*

$$C^*(\theta^*, \delta) \sim \mathcal{N}_{2m}(\mathbf{0}, \Omega(\theta_0)).$$

*Then the  $\alpha$ -ECS test for the problem  $\mathbf{H}_0 : \theta^* = \theta_0$  vs.  $\mathbf{H}_1 : \theta^* \neq \theta_0$  in the limiting experiment of a weakly identified GMM model rejects the null hypothesis if*

$$m(\theta_0)'m(\theta_0) > \chi_{m,1-\alpha}^2.$$

See Appendix A.3.9 for a proof of Result 3.

The test, evaluated at sample analogues, coincides with the  $S$ -test of Stock and Wright

(2000):

$$\left(\frac{1}{T} \sum_{t=1}^T h(x_t, \theta_0)\right)' [\Omega(\theta_0)]_m^{-1} \left(\frac{1}{T} \sum_{t=1}^T h(x_t, \theta_0)\right) > \chi_{m,1-\alpha}^2,$$

where  $[\Omega(\theta_0)]_m$  is the asymptotic variance of

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T h(x_t, \theta_0),$$

and, to simplify notation, I have assumed that such a covariance matrix is known.<sup>24</sup>

EXAMPLE (GMM-IV): The condition of Result 3 is simple to verify in the GMM-IV model. Note that  $k+1 \geq 2k$  if and only if  $k=1$ . In this case, the  $S$ -test is ECS and rejects if:

$$\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T Z_t(y_t - \theta_0 Y_t)\right)' \left(\frac{1}{T} \sum_{t=1}^T Z_t Z_t' (y_t - \theta_0 Y_t)^2\right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T Z_t(y_t - \theta_0 Y_t)\right) > \chi_{1,1-\alpha}^2,$$

which coincides with the robust version of the AR test derived in section 5.1. When  $k > 1$ , there are no priors over  $(\theta^*, \delta)$  for which the function  $C^*(\theta^*, \delta)$  behaves as a Gaussian distribution on  $\mathbb{R}^{2k}$ . This is simply because  $C^*$  takes values on a strict subset of  $\mathbb{R}^{2k}$ .

Now, I will provide a general expression for ECS tests in weakly identified GMM models.

Let

$$\gamma(\theta_0)' = [m(\theta_0)', d(\theta_0)'].$$

**Result 4.** *Let  $p_1$  be a full-support prior over  $\mathbb{R}^{n+1}$  and suppose that assumptions*

---

<sup>24</sup>If this were not the case, one could replace  $\Omega(\theta_0)$  with an estimator  $\hat{\Omega}_T(\theta_0)$ . The (Gaussian) weak convergence assumption combined with  $\hat{\Omega}_T(\theta_0) \xrightarrow{P} \Omega(\theta_0)$  yield the same limiting experiment.

RGMM1 and R2 hold. Then the  $\alpha$ -ECS test statistic for the problem  $\mathbf{H}_0 : \theta^* = \theta_0$  vs.  $\mathbf{H}_1 : \theta^* \neq \theta_0$  rejects the null hypothesis if

$$\int_{\mathbb{R}^{n+1}} \exp\left(\gamma(\theta_0)' D(\theta_0) C^*(\theta^*, \delta)\right) \exp\left(-\frac{1}{2} C^*(\theta^*, \delta)' \Omega(\theta_0)^{-1} C^*(\theta^*, \delta)\right) p_1(\theta^*, \delta) d\theta^* d\delta,$$

is larger than its  $1 - \alpha$  quantile, conditional on  $d(\theta_0)$ .

See Appendix A.3.10 for a proof of Result 4.

EXAMPLE (GMM-IV): The sample analogue of the test statistic in Result 4 is given by

$$\begin{aligned} z(\widehat{m}(\theta_0), \widehat{d}(\theta_0)) &= \int_{\mathbb{R}^{k+1}} \left[ \exp\left(\begin{pmatrix} \widehat{m}(\theta_0) \\ \widehat{d}(\theta_0) \end{pmatrix}' \widehat{D}(\theta_0) \begin{pmatrix} (\theta^* - \theta_0)\delta \\ \delta \end{pmatrix}\right) \right. \\ &\quad \left. \exp\left(-\frac{1}{2} \begin{pmatrix} (\theta^* - \theta_0)\delta \\ \delta \end{pmatrix}' \widehat{\Omega}(\theta_0)^{-1} \begin{pmatrix} (\theta^* - \theta_0)\delta \\ \delta \end{pmatrix}\right) p_1(\theta^*, \delta)\right] d\theta^* d\delta, \end{aligned}$$

where

$$\widehat{D}(\theta_0) = \begin{pmatrix} \left[\frac{1}{T} \sum_{t=1}^T Z_t Z_t' (y_t - \theta_0 Y_t)^2\right]^{-1/2} & \mathbf{0} \\ \widehat{d}_1 & \widehat{d}_2 \end{pmatrix}, \quad \widehat{\Omega}(\theta_0)^{-1} = \widehat{D}(\theta_0)' \widehat{D}(\theta_0),$$

and  $\widehat{d}_1, \widehat{d}_2$  are the sample analogues of the matrix defined in Lemma 3 in Appendix A applied to  $\Omega(\theta_0)$ . The boundary sufficient statistic is given by:

$$\widehat{d}(\theta_0) = \widehat{d}_1 \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_t (y_t - \theta_0 Y_t) + \widehat{d}_2 \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_t Y_t,$$

and

$$\widehat{m}(\theta_0) = \left( \frac{1}{T} \sum_{t=1}^T Z_t Z_t' (y_t - \theta_0 Y_t)^2 \right)^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^T Z_t (y_t - \theta_0 Y_t).$$

The critical value  $c(\widehat{d}(\theta_0))$  is obtained by fixing  $\widehat{d}(\theta_0)$  and computing the  $1-\alpha$  quantile of the random variable:

$$z(m, \widehat{d}(\theta_0)), \quad m \sim \mathcal{N}_k(\mathbf{0}, \mathbb{I}_k).$$

The ECS test rejects the null hypothesis if  $z(\widehat{m}(\theta_0), \widehat{d}(\theta_0)) > c(\widehat{d}(\theta_0))$ .

### 3.5.3 Dynamic Effects in Structural VARs

Montiel, Stock, and Watson (2012) develop methods for inference in structural vector autoregressions (SVARs) in which the structural shocks are identified using external instruments. They focus on the possibility of potentially weak instruments. This section shows that the test used by Montiel, Stock, and Watson (2012) to build confidence intervals for dynamic effects in “just-identified” SVARs is ECS. The test rejects for large values of the sample covariance between the instrument that identifies the target structural shock and a linear combination of all the reduced-form shocks in the vector autoregression. This section also presents an ECS test for the over-identified SVAR model.

**a) ECONOMETRIC MODEL:** Let the  $r \times 1$  time series  $Y_t$  follow the reduced-form stationary VAR with  $p$  lags:

$$Y_t = A(L)Y_{t-1} + \eta_t,$$

where  $A(L)$  is a known lag-polynomial that is assumed invertible. The  $r \times 1$  vector  $\eta_t$

represents the reduced form innovations.<sup>25</sup> There is a  $r \times 1$  vector of structural shocks  $\varepsilon_t$  that satisfy:

$$\eta_t = H\varepsilon_t = [H_1, H_2, \dots, H_r]\varepsilon_t.$$

The unknown  $r \times r$  matrix  $H$  is assumed invertible,  $H_i$  denotes the  $i$ -th column of  $H$  and  $h_{im}$  denotes the  $m$ -th element of  $H_i$ . The structural moving average representation of the reduced-form VAR is given by:

$$Y_t = A(L)^{-1}H\varepsilon_t.$$

Let  $C'_{hj} = (c_{hj1}, c_{hj2}, \dots, c_{hjr})$  denote the  $j$ -th row of the  $h$ -lag matrix of  $A(L)^{-1}$ . The object of interest is the dynamic effect of a shock  $\varepsilon_{1t}$  over variable  $j$  at horizon  $h$ . The null hypothesis states that an impulse in the structural shock  $\varepsilon_{1t}$  will have an effect of  $\kappa_0$  over the  $j$ -th component of  $Y_{t+h}$ ; that is

$$\mathbf{H}_0 : C'_{hj}H_1 = \kappa_0 \quad vs. \quad \mathbf{H}_1 : C'_{hj}H_1 \neq \kappa_0.$$

**b) DISTRIBUTIONAL ASSUMPTIONS:** The  $k \times 1$  vector of external instruments  $Z_t$  is used to identify the dynamic effect with respect to the structural shock of interest through the moment condition

$$\mathbb{E}[\varepsilon_t \otimes Z_t] = e_1 \otimes \alpha, \quad e_1 \in \mathbb{R}^r, \quad e_1 = (1, 0, 0, \dots, 0)'; \quad \alpha \in \mathbb{R}^k,$$

which implies that

$$\mathbb{E}[\eta_t \otimes Z_t] = \mathbb{E}[H\varepsilon_t \otimes Z_t] = (H \otimes I_k)\mathbb{E}[\varepsilon_t \otimes Z_t] = H_1 \otimes \alpha.<sup>26</sup>$$

---

<sup>25</sup>Assuming  $A(L)$  is known entails no loss of generality; see Montiel, Stock and Watson (2012) for details.

<sup>26</sup>Note that  $\{\eta_t\}$  is observed as  $A(L)$  is assumed known.

In order to “normalize” the effect of interest, it is assumed that  $h_{11} = 1$ .<sup>27</sup> Consequently, the moment condition above identifies  $H_1$  and the parameter of interest,  $C'_{hj}H_1$ .

To model the potential weak correlation between the instruments  $(Z_t)$  and the structural shock  $(\varepsilon_{1t})$ , assume that  $\alpha = a/\sqrt{T}$ .

**c) STATISTICAL MODEL:** The data generating process for  $\{\eta_t, Z_t\}_{t=1}^T$  is restricted by imposing the following weak convergence assumptions

$$(1/\sqrt{T}) \sum_{t=1}^T \eta_t \otimes Z_t = \begin{pmatrix} (1/\sqrt{T}) \sum_{t=1}^T \eta_{1t} Z_t \\ (1/\sqrt{T}) \sum_{t=1}^T \eta_{\bullet t} \otimes Z_t \end{pmatrix} \xrightarrow{d} \gamma \sim \mathcal{N}_{rk}(H_1 \otimes a, \Omega)$$

where

$$\eta_{\bullet t} = \begin{pmatrix} \eta_{2t} \\ \vdots \\ \eta_{rt} \end{pmatrix}.$$

**d) BOUNDARY SUFFICIENCY:** Consider first the just-identified case ( $k = 1$ ). Let  $\kappa \equiv C'_{hj}H_1$  and let  $\kappa_0$  denote the null hypothesis of interest. Define

$$C'_0 = (c_{hj1} - \kappa_0, c_{hj2}, \dots, c_{hjr}),$$

and let  $C_0^{\perp'}$  denote the *orthonormal part* of  $C'_0$ ; that is,  $C_0^{\perp'}$  is the  $(r-1) \times r$  matrix such that

$$(i) \text{ The } r \times r \text{ matrix } \begin{pmatrix} C'_0 \\ C_0^{\perp'} \end{pmatrix} \text{ has full rank.}$$

---

<sup>27</sup>Therefore, the structural shock  $h_{11}$  is measured in units of the variable  $y_{1t}$  in the reduced form VAR.



(ii)  $C_0' C_0^\perp = \mathbf{0}$ .

(iii) For each  $m = 1 \dots r - 1$ ,  $C_{0m}^{\perp'} C_{0m}^\perp = 1$ ; where  $C_{0m}^{\perp'}$  denotes the  $m$ -th row of  $C_0^{\perp'}$ .

The standardized limiting experiment for the SVARs testing problem is given by

$$\begin{pmatrix} S \\ T_1 \end{pmatrix} \equiv \begin{pmatrix} C_0' \gamma / (C_0' \Omega C_0)^{1/2} \\ (C_0^{\perp'} \Omega^{-1} C_0^\perp)^{-1/2} C_0^{\perp'} \Omega^{-1} \gamma \end{pmatrix} \\ \sim \mathcal{N}_r \left( \begin{pmatrix} (\kappa - \kappa_0) a \\ (C_0^{\perp'} \Omega^{-1} C_0^\perp)^{-1/2} C_0^{\perp'} \Omega^{-1} H_1 a \end{pmatrix}, \mathbb{I}_r \right).$$

Under the null hypothesis  $\kappa - \kappa_0 = 0$ . Therefore,  $T_1$  is a boundedly complete boundary-sufficient statistic, whenever  $k = 1$ .

**e1) PRIORS FOR JUST-IDENTIFIED SVARs:** Let

$$\begin{pmatrix} m_1 \\ \vdots \\ m_r \end{pmatrix} \sim \mathcal{N}_r(\mathbf{0}, \Omega),$$

and consider the following distribution over the parameters  $(H_1, a)$  of the statistical model above

$$a = m_1, \quad h_{1n} = m_n / m_1, \quad n = 2, \dots, r.$$

**Result 5.** *The  $\alpha$ -ECS test for the problem  $\mathbf{H}_0 : \kappa = \kappa_0$  vs.  $\mathbf{H}_1 : \kappa \neq \kappa_0$  in a just-*

identified SVAR model with priors in **e1**) rejects the null hypothesis if:

$$S'S > \chi_{1,1-\alpha}^2.$$

See Appendix A.3.11 for a proof of Result 5.

The sample analogue of the ECS test statistic above is

$$\left( \frac{1}{\sqrt{T}} \sum_{t=1}^T (C_0' \eta_t) Z_t \right)^2 / C_0' \Omega C_0,$$

which is equivalent to the “robust” AR statistic (Result 1\*) from the following just-identified IV model:

- a) Outcome variable:  $C_{hj}' \eta_t$
- b) Endogenous variable:  $\eta_{1t}$  (with coefficient  $\kappa$ )
- c) Instrument:  $Z_t$
- d) Null hypothesis:  $\kappa = \kappa_0$

**Remark 15.**  $C_0$  and  $\eta_t$  have been assumed to be known sample objects.

**e2) PRIORS FOR THE OVER-IDENTIFIED MODEL WITH A “KRONECKER” COVARIANCE MATRIX:** Suppose  $\Omega = \Sigma \otimes Q = \mathbb{E}[\eta_t \eta_t'] \otimes \mathbb{E}[Z_t Z_t']$ . In Appendix A.3.11 it is shown that the limiting experiment of the SVARs testing problem with a kronecker covariance matrix admits the following representation

$$\begin{pmatrix} S \\ T_1 \\ \vdots \\ T_{r-1} \end{pmatrix} \sim \mathcal{N}_{rk}(\phi \otimes \rho\omega, \mathbb{I}_r \otimes \mathbb{I}_k).$$

The sample space is  $R^{rk}$  with a typical element denoted by  $(S', T'_1, T'_2 \dots T'_{r-1})'$ . The parameter space is given by  $\mathbb{R}_+ \times \mathcal{S}_R^{r-1} \times \mathcal{S}^{k-1}$ , with typical element  $(\rho, \phi, \omega)$ .<sup>28</sup> The hypothesis  $\kappa = \kappa_0$  is equivalent to:

$$\mathbf{H}_0 : \phi_1 = 0 \quad vs. \quad \mathbf{H}_1 : \phi_1 \neq 0.$$

Hence,  $T \equiv (T'_1, T'_2, \dots T'_{r-1})'$  is a boundary-sufficient statistic. Consider the independent priors

$$\rho \sim \sqrt{\lambda^2 \chi_k^2}, \quad \omega \sim \mathcal{U}(\mathcal{S}^k), \quad \phi \sim \mathcal{U}(\mathcal{S}_R^{r-1})$$

$$\text{Let } M \equiv [S, T_1, \dots T_{r-1}][S, T_1, \dots T_{r-1}]'.$$

**Result 6.** *The ECS test for the problem  $\mathbf{H}_0 : \phi_1 = 0$  vs.  $\mathbf{H}_1 : \phi_1 \neq 0$  in the over-identified SVAR model with priors in e2) rejects the null hypothesis if the statistic*

$$\int_{\mathcal{S}_R^{r-1}} \exp\left(\frac{\lambda^2}{2(1+\lambda^2)} \phi' M \phi\right) d\lambda_{\mathcal{S}^{r-1}}(d\phi)$$

*is larger than its  $1 - \alpha$  quantile, conditional on  $T$ .*

See Appendix A.3.12 for a proof of Result 6.

The sample analogues of  $S$  and  $T$  are given by

$$\hat{S} \equiv (1/\sqrt{T}) \sum_{t=1}^T \left( C'_0 \eta_t / (C'_0 \hat{\Sigma} C'_0)^{1/2} \otimes \hat{Q}^{-1/2} Z_t \right),$$

---

<sup>28</sup>  $\mathcal{S}_R^{r-1} = \{\phi \in \mathcal{S}^{r-1} \mid e_1 \Sigma A' \phi \geq 0\}$ .

and

$$\hat{T} = (1/\sqrt{T}) \sum_{t=1}^T \left( \left( C_0^{\perp'} \hat{\Sigma}^{-1} C_0^{\perp} \right)^{-1/2} C_0^{\perp'} \hat{\Sigma}^{-1} \eta_t \otimes \hat{Q}^{-1/2} Z_t \right).$$

### 3.6 Summary of the main results and Conclusion

Boundary sufficiency arises naturally in three widely used models in econometrics: Linear Instrumental Variables Regression (IV), the Generalized Method of Moments (GMM), and Structural Vector Autoregressions (SVARs). This property is common to other hypothesis testing problems with nuisance parameters; for example, the Linear Regression Model with a sign restriction in Elliott *et al.* (2012); the predictive regression model with nearly integrated regressors studied in Stock and Watson (1996), Jansson and Moreira (2006), and Elliott *et al.* (2012); and testing problems in exponential family models. Boundary sufficiency is an attractive feature, for it allows the econometrician to control the rate of Type I error—which can dramatically vary in standard Wald tests—regardless of the values of nuisance parameters.

This paper used statistical decision theory as a guiding principle to derive a new class of tests for hypothesis testing problems with a boundary-sufficient statistic. The tests are *efficient*, for they minimize a weighted sum of the average rates of Type I and Type II error (average risk). The tests are *conditionally similar on the boundary*, for they control the rate of Type I error on the set of null parameter values that are the closest to the alternative set by conditioning on the realizations of the boundary-sufficient statistic.

This paper showed that *Efficient Conditionally Similar-on-the-boundary* (ECS) tests are *admissible* within the class of conditionally similar-on-the-boundary procedures. Moreover, ECS tests were shown to verify an important finite-sample optimality property: admissibility within the class of *all* tests, provided the boundary-sufficient statis-

tic is boundedly complete.

Theorem 1 in this paper provided a systematic method to derive admissible tests within the class of conditionally and unconditionally similar-on-the-boundary tests. The idea is conceptually simple: it suffices to trade off the average rates of Type I and Type II error using a monotone continuous function  $\mathbf{W} : \mathbb{R}^2 \rightarrow \mathbb{R}$ . When  $\mathbf{W}$  is linear, the exercise is equivalent to average risk minimization (using full-support priors) subject to a conditional or unconditional similarity-on-the-boundary constraint. The solution to this problem is well-defined, for the domain of the optimization problem is *weak\* compact* and the objective function is *weak\* continuous*. ECS tests are thus defined as the solution to a minimization problem over a space of functions.

Theorem 2 in this paper showed that the minimization problem defining ECS tests has a convenient closed form solution: the ECS test statistic can be expressed as a linear combination of the null and alternative integrated likelihoods. The critical value *function* is given by the conditional quantiles of the ECS test statistic. Theorem 1 and 2 complement recent findings by Moreira and Moreira (2012), which develop methods to approximate the solution of risk minimization problems over the space of similar tests, without requiring a boundary-sufficient statistic.

This paper applied the theory of ECS tests to hypothesis testing problems in IV, GMM, and SVARs. The emphasis was on “point” problems, in which case the ECS tests maximize weighted average power (WAP) subject to a conditional similarity constraint on the null set.

Result 1 showed that the Anderson and Rubin (1949) (AR) test is ECS in just-identified IV models with Gaussian reduced-form errors, independent observations, fixed instruments, and an arbitrary number of endogenous regressors. Result 1\* extended this result to models with heteroskedastic, autocorrelated, and/or clustered data. The priors over the structural parameters of the IV model ( $\beta$  and  $\Pi$ ) for which

the AR test maximizes WAP were shown to have an interesting feature: there are no other priors for which the implied distribution over the reduced-form parameters ( $\Pi\beta$  and  $\Pi$ ) is Gaussian, centered at zero, and with the same covariance matrix as the distribution of their sample counterparts. Since the boundary-sufficient statistic in the just-identified IV model is boundedly complete (regardless of the number of endogenous regressors), the AR is admissible in the class of all tests.

Result 2 derived a novel test for point-null hypotheses in the over-identified IV model studied by Andrews *et al.* (2006) and Chamberlain (2007). The new ECS test enjoys basic optimality properties that neither CLR nor the TSLS (LIML) Wald tests have been shown to satisfy: namely, admissibility in the class of all tests and efficiency in the class of similar tests. The “conditional” critical region of the new test can be expressed in terms of the AR and the Lagrange Multiplier (LM) statistics. If the LM is below (above) its conventional  $\chi_1^2$  critical value, the ECS test automatically adjusts upwards (downwards) the  $\chi_k^2$  threshold for the AR. The magnitude of the adjustment depends on the value of the boundary-sufficient statistic and the ECS test rejects the null hypothesis whenever the AR exceeds the adjusted critical value. Result 2\* derived a new test for one-sided problems.

Result 3 showed that the  $S$ -test of Stock and Wright (2000) is ECS in some weakly identified GMM models in which the parameter of interest is scalar and the population moment function is known up to a finite-dimensional vector. To the best of my knowledge, this is the first optimality result derived for weakly identified GMM models. A key component of Result 3 is the theory of Müller (2011), which motivates the study of a statistical model derived from a set of weak convergence assumptions—as opposed to assumptions about the finite distribution of the data.

Result 4 provided a general expression for ECS tests in GMM models. An important observation is that the boundary-sufficient statistic has the same dimension as

the derivative of the sample moment condition. In general, it is not clear whether a *dimension reduction* for the boundary-sufficient statistic is available—as it is assumed by Kleibergen (2007), whose tests condition on a scalar statistic containing information about the rank of the matrix of derivatives.

Finally, Result 5 showed that the test used by Montiel, Stock, and Watson (2012) to build confidence intervals for dynamic effects in “just-identified” SVARs is ECS. The test rejects for large values of the sample covariance between the instrument that identifies the target structural shock in the model and a linear combination of all the reduced-form shocks in the vector autoregression. Result 6 derived ECS tests for the over-identified SVAR model.

There are two observations that highlight the importance of *Efficient Conditionally Similar-on-the-boundary* (ECS) tests.

First, efficient tests and, more generally, admissible tests need not be (conditionally) similar. For instance, standard Bayes tests—that is, tests that reject for large values of a ratio of integrated likelihoods—are, by construction, efficient and thus admissible. Despite their admissibility, Bayes tests face an important limitation: their rate of Type I error can vary over the null set and, in some cases, such rate can be arbitrarily close to one regardless of the critical value used in their implementation.

Second, there are already tests in the literature that use a boundary-sufficient statistic to control the rate of Type I error. For example, Moreira (2003) proposed the Conditional Likelihood Ratio test (CLR) for IV and Kleibergen (2007) extended the CLR to GMM problems. These procedures are, by construction, conditionally similar on the boundary. However, as far as I know, they are neither admissible nor efficient, even in some restricted sense. In fact, it is not clear whether the CLR and its extensions are admissible within the class of conditionally similar tests. Without an analytical claim for admissibility, there is no guarantee that these procedures cannot be improved.

**CONCLUDING REMARK:** A continued focus of the econometrics literature in the past two decades has been the *finite-sample* analysis of widely used estimation and testing procedures. Studying the performance of any statistical decision rule (e.g., estimator, test, or confidence interval) in a finite sample requires—in one way or another—that *there be a statistical model*, which inevitably connects the properties of the decision rules under consideration with *statistical decision theory*. It is then possible to use classical concepts—for example, that of a risk function—to think about *optimal* selection of estimators, tests, or confidence intervals. This paper followed this approach. The decision problem of interest was hypothesis testing in IV, GMM, and SVARs. The main optimality concept was that of finite-sample admissibility. There was an additional constraint motivated by applied work: invariance of the rates of Type I error with respect to nuisance parameters—i.e., similarity—in some region of the null set. This paper identified a common statistical property in the three problems under consideration: boundary sufficiency. This property was used to derive a new class of tests and to establish a new sense of efficiency for some existing procedures.



# References

- ABRAMOWITZ, M. and STEGUN, I. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55. Dover publications.
- ALIPRANTIS, C. and BORDER, K. (2006). *Infinite dimensional analysis: a hitchhiker's guide*. Springer Verlag.
- ANDERSON, T. and RUBIN, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, **20** (1), 46–63.
- ANDREWS, D., MOREIRA, M. and STOCK, J. (2006). Optimal two-sided invariant similar tests for instrumental variables regression.”. *Econometrica*, **74** (3), 715–52.
- and STOCK, J. (2006). Inference with weak instruments. *Advances in Economics and Econometrics, Theory and Applications (9th Congress of the Econometric Society)*, **3** (Chapter 6), Cambridge University Press.
- ANTOINE, B. and LAVERGNE, P. (2012). Conditional moment models under semi-strong identification. *Working Paper, Simon Fraser University*.
- ATTEMA, A., BLEICHRODT, H., ROHDE, K. and WAKKER, P. (2010). Time-Tradeoff Sequences for Analyzing Discounting and Time Inconsistency. *Management Science*, **56** (11), 2015–2030.
- BANSAL, R. and YARON, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance*, **59**, 1481–1509.
- BAUM, C. F., SCHAFFER, M. E. and STILLMAN, S. (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *The Stata Journal*, **7** (4), 465–506.
- BERGER, J. (1985). *Statistical decision theory and Bayesian analysis*. Springer.
- BILLINGSLEY, P. (1995). Probability and measure. *John Wiley & Sons*.
- BLEICHRODT, H., ROHDE, K. and WAKKER, P. (2008). Koopmans’ constant discounting for intertemporal choice: A simplification and a generalization. *Journal of Mathematical Psychology*, **52** (6), 341–347.

- BOCK, M., JUDGE, G. and YANCEY, T. (1984). A simple form for the inverse moments of non-central  $\chi^2$  and  $f$  random variables and certain confluent hypergeometric functions. *Journal of Econometrics*, **25** (217-234).
- BOUND, J., JAEGER, D. and BAKER, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, **90** (430).
- BUN, M. and DE HAAN, M. (2010). Weak instruments and the first stage  $f$ -statistic in iv models with a nonscalar error covariance structure. *University of Amsterdam*.
- CAMPBELL, J. (2003). Consumption-based asset pricing. *Handbook of the Economics of Finance*, **1**, 803–887.
- and MANKIW, N. (1989). Consumption, income and interest rates: Reinterpreting the time series evidence.
- CASARI, M. and DRAGONE, D. (2010). On negative time preferences. *Economics Letters*.
- CHAMBERLAIN, G. (2007). Decision theory applied to an instrumental variables model. *Econometrica*, **75** (3), 609–652.
- CHAO, N. R. S. J. A. H., JOHN C. and NEWEY, W. K. (2012). Asymptotic distribution of  $J$  in the heteroskedastic iv regression with many instruments. *Econometric Theory*, **28**, 42–86.
- CHERNOZHUKOV, V., HANSEN, C. and JANSSON, M. (2009). Admissible invariant similar tests for instrumental variables regression. *Econometric Theory*, **25** (03), 806–818.
- CHEW, H. and KARNI, E. (1994). Choquet expected utility with a finite state space: Commutativity and act-independence. *Journal of Economic Theory*, **62** (2), 469–479.
- CONERLY, M. D. and MANSFIELD, E. R. (1988). An approximate test for comparing heteroscedastic regression models. *Journal of the American Statistical Association*, **83** (403), 811–817.
- DELLAVIGNA, S. and MALMENDIER, U. (2004). Contract Design And Self-Control: Theory And Evidence. *Quarterly Journal of Economics*, **119** (2), 353–402.
- DIAMOND, P. and KÖSZEGI, B. (2003). Quasi-hyperbolic discounting and retirement. *Journal of Public Economics*, **87** (9-10), 1839–1872.
- ELLIOTT, G., MÜLLER, U. and WATSON, M. (2012). Nearly optimal tests when a nuisance parameter is present under the null hypothesis.

- EPSTEIN, L. and ZIN, S. (1989a). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica: Journal of the Econometric Society*, pp. 937–969.
- and — (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy*, pp. 263–286.
- EPSTEIN, L. G. and ZIN, S. (1989b). Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica*, **57** (4), 937–969.
- FAREBROTHER, R. (1990). Algorithm as 256: The distribution of a quadratic form in normal variables. *Applied Statistics*, **39** (2), 294–309.
- FERGUSON, T. (1967). *Mathematical statistics: A decision theoretic approach*, vol. 7. Academic Press New York.
- FISHBURN, P. and RUBINSTEIN, A. (1982). Time preference. *International Economic Review*, **23** (3), 677–694.
- FUDENBERG, D. and MASKIN, E. (1991). On the dispensability of public randomization in discounted repeated games. *Journal of Economic Theory*, **53** (2), 428–438.
- GHIRARDATO, P. and MARINACCI, M. (2001). Risk, ambiguity, and the separation of utility and beliefs. *Mathematics of Operations Research*, pp. 864–890.
- GHOSH, B. (1973). Some monotonicity theorems for  $\chi^2$ , f and t distributions with applications. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 480–492.
- GORMAN, W. (1968). The structure of utility functions. *The Review of Economic Studies*, **35** (4), 367–390.
- GRUBBS, F. E. (1964). Approximate circular and noncircular offset probabilities of hitting. *Operations Research*, **12** (1), 51–62.
- GUGGENBERGER, P. (2010a). The impact of a hausman pretest on the asymptotic size of a hypothesis test. *Econometric Theory*, **26**, 369–382.
- (2010b). The impact of a hausman pretest on the size of a hypothesis test: The panel data case. *Journal of Econometrics*, **156**, 337–343.
- GUL, F. and PESENDORFER, W. (2001). Temptation and self-control. *Econometrica*, pp. 1403–1435.
- HALL, R. E. (1988). Intertemporal substitution in consumption. *The Journal of Political Economy*, **96** (2), 339–357.

- HANSEN, L., HEATON, J. and YARON, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, **14** (3), 262–280.
- and SINGLETON, K. (1983). Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *The Journal of Political Economy*, pp. 249–265.
- HAUSMAN, J. A., NEWEY, W. K., WOUTERSEN, T. and CHAO, J. C. (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics*, **3**, 211–255.
- HAYASHI, T. (2003). Quasi-stationary cardinal utility and present bias. *Journal of Economic Theory*, **112** (2), 343–352.
- HENSHAW, R. C. (1966). Testing single-equation least squares regression models for autocorrelated disturbances. *Econometrica*, **34** (3), 646–660.
- IMHOF, J. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, **48** (3-4), 419.
- JANSSON, M. and MOREIRA, M. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica: Journal of the Econometric Society*, **74** (3), 681–714.
- JOHNSON, N., KOTZ, S. and BALAKRISHNAN, N. (1995). Continuous univariate distributions, vol. 2. 1995.
- KLEIBERGEN, F. (2007). Generalizing weak instrument robust iv statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, **139** (1), 181–216.
- KOERTS, J. and ABRAHAMSE, A. (1969). *On the theory and application of the general linear model*. Rotterdam University Press.
- KOOPMANS, T. (1960). Stationary ordinal utility and impatience. *Econometrica: Journal of the Econometric Society*, **28** (2), 287–309.
- (1972). Representation of preference orderings over time. In C. McGuire and R. Radner (eds.), *Decision and organization*, pp. 79–100.
- KREPS, D. and PORTEUS, E. (1978). Temporal Resolution of Uncertainty and Dynamic Choice Theory. *Econometrica*, **46** (1), 185–200.
- LAIBSON, D. (1997). Golden Eggs and Hyperbolic Discounting\*. *Quarterly Journal of Economics*, **112** (2), 443–477.
- (1998). Life-cycle consumption and hyperbolic discount functions. *European Economic Review*, **42** (3-5), 861–871.

- , REPETTO, A. and TOBACMAN, J. (2007). Estimating discount functions with consumption choices over the lifecycle. *NBER working paper*.
- LE CAM, L. (1986). *Asymptotic methods in statistical decision theory*. Springer.
- LEEB, H. and POETSCHER, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, **21**, 21–59.
- LEHMANN, E. and ROMANO, J. (2005). *Testing statistical hypotheses*. Springer Verlag.
- LINNIK, J. (1968). *Statistical problems with nuisance parameters*, vol. 20. American Mathematical Society.
- LIU, H., TANG, Y. and ZHANG, H. (2009). A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Computational Statistics & Data Analysis*, **53** (4), 853–856.
- LIU, S. and NEUDECKER, H. (1995). Matrix-trace cauchy-schwarz inequalities and applications in canonical correlation analysis. *Statistical Papers*, **36** (1), 287–298.
- LOEWENSTEIN, G. and PRELEC, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, **107** (2), 573–597.
- LUCE, R. D. and TUKEY, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, (1), 1–27.
- MARDIA, K. and JUPP, P. (2000). *Directional statistics*, vol. 28. Wiley.
- MAVROEIDIS, S., PLAGBORG-MOLLER, M. and STOCK, J. A. (2012). Empirical evidence on inflation expectations in the new keynesian phillips curve. Working Paper.
- MCCLURE, S., ERICSON, K., LAIBSON, D., LOEWENSTEIN, G. and COHEN, J. (2007). Time discounting for primary rewards. *Journal of Neuroscience*, **27** (21), 5796.
- MIKUSHEVA, A. (2010). Robust confidence sets in the presence of weak instruments. *Journal of Econometrics*.
- MONTIEL OLEA, J. and PFLUEGER, C. (2012). A robust test for weak instrumets. Working Paper.
- , STOCK, J. and WATSON, M. (2012). Inference in structural vars with external instruments. *Working Paper*.
- MOREIRA, M. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, **71** (4), 1027–1048.
- (2009). Tests with correct size when instruments can be arbitrarily weak. *Journal of Econometrics*, **152** (2), 131–140.

- and MOREIRA, H. (2012). Contributions to the theory of similar tests. Working Paper.
- MÜLLER, U. K. (2011). Efficient tests under a weak convergence assumption. *Econometrica: Journal of the Econometric Society*, **79** (2), 395–435.
- MUNKRES, J. (2000). *Topology (Second Edition)*. Prentice Hall Inc.
- NAGAR, A. (1959). The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica*, **27** (4), 575–595.
- NAKAMURA, Y. (1990). Subjective expected utility with non-additive probabilities on finite state spaces. *Journal of Economic Theory*, **51** (2), 346–366.
- NELSON, C. and STARTZ, R. (1990). The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *Journal of Business*, **63** (1), 125–140.
- NEYMAN, J. (1935). Sur la vérification des hypothèses statistiques composées. *Bull. Soc. Math. France*, (63), 346–366.
- NOOR, J. (2009). Hyperbolic discounting and the standard model: Eliciting discount functions. *Journal of Economic Theory*, **144** (5), 2077–2083.
- (2011). Temptation and revealed preference. *Econometrica*, **79** (2), 601–644.
- PATNAIK, P. (1949). The non-central chi-square and f-distributions and their applications. *Biometrika*, **36** (1-2), 202.
- PEARSON, E. (1959). Note on an approximation to the distribution of non-central chi-squared. *Biometrika*, **46** (3/4), 364.
- PHELPS, E. and POLLAK, R. (1968). On second-best national saving and game-equilibrium growth. *The Review of Economic Studies*, pp. 185–199.
- RAMSEY, F. (1926). Truth and probability. In D. H. Mellor (ed.), *Philosophical Papers*, Cambridge: Cambridge University Press, 1990.
- ROTHENBERG, T. (1984). Approximating the distributions of econometric estimators and test statistics. *Handbook of econometrics*, **2**, 881–935.
- RUDIN, W. (2005). Functional analysis. international series in pure and applied mathematics.
- (2006). *Real and complex analysis*. Tata McGraw-Hill Education.
- RUUD, P. (2000). *An introduction to classical econometric theory*. Oxford University Press New York.

- SAMUELSON, P. (1937). A note on measurement of utility. *The Review of Economic Studies*, **4** (2), 155–161.
- SARGAN, J. (1974). The validity of nagar’s expansion for the moments of econometric estimators. *Econometrica: Journal of the Econometric Society*, pp. 169–176.
- SATTERTHWAITE, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, **2** (6), 110–114.
- SORIN, S. (1986). On repeated games with complete information. *Mathematics of Operations Research*, **11** (1), 147–160.
- STAIGER, D. and STOCK, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, **65** (3), 557–586.
- STEIN, E. (2011). *Functional Analysis: Introduction to Further Topics in Analysis*. Princeton Univ Pr.
- STOCK, J. and WATSON, M. (1996). Confidence sets in regressions with highly serially correlated regressors. *manuscript, Harvard University*.
- and WRIGHT, J. (2000). Gmm with weak identification. *Econometrica*, **68** (5), 1055–1096.
- , — and YOGO, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, **20** (4), 518–529.
- and YOGO, M. (2005). Testing for weak instruments in linear iv regression, identification and inference for econometric models: Essays in honor of thomas rothenberg. *ed. D.W. Andrews and J.H. Stock*, pp. 80–108.
- STROOCK, D. (1999). *A concise introduction to the theory of integration*. Birkhauser.
- SZEKELY, G. and BAKIROV, N. (2003). Extremal probabilities for gaussian quadratic forms. *Probability Theory and Related Fields*, **126** (2), 184–202.
- THALER, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, **8** (3), 201–207.
- THEIL, H. and NAGAR, A. (1961). Testing the independence of regression disturbances. *Journal of the American Statistical Association*, **56** (296), 793–806.
- VAN DER VAART, A. (2000). *Asymptotic statistics*. 3, Cambridge Univ Pr.
- VIND, K. (1971). Note on “The Structure of Utility Functions”. *The Review of Economic Studies*, **38** (1), 113–113.

- WALD, A. (1950). Statistical decision functions. *New York*, **660**.
- WHITE, H. (1980). Nonlinear regression on cross-section data. *Econometrica: Journal of the Econometric Society*, pp. 721–746.
- (2001). *Asymptotic theory for econometricians*. Academic press New York.
- WOOLDRIDGE, J. (2002). *Econometric analysis of cross section and panel data*. The MIT press.
- YEH, J. (2000). *Lectures on Real Analysis*. World Scientific, Singapore.
- YOGO, M. (2004). Estimating the elasticity of intertemporal substitution when instruments are weak. *Review of Economics and Statistics*, **86** (3), 797–810.
- ZHAN, Z. (2010). Detecting weak identification by bootstrap. *Job Market Paper*.
- ZHANG, F. (2010). *Matrix Theory: Basic Results and Techniques, Second Edition*. Springer.



# Appendix A

## Proofs

### A.1 Proofs for Chapter 1

#### A.1.1 Proof of Theorem 1

Necessity of the axioms is straightforward. For sufficiency, we follow a sequence of steps.

**Step 1.** The initial separability axiom guarantees that the sets  $\{0, 1\}$ ,  $\{1, 2\}$ , and  $\{1, 2, \dots\}$  are independent. To show that for all  $t = 2, \dots$  the sets  $\{t, t + 1\}$  are independent fix  $x, y, z, z' \in \mathcal{F}$  and suppose that

$$(z_0, z_1, \dots, z_{t-1}, x_t, x_{t+1}, z_{t+1}, \dots) \succ (z_0, z_1, \dots, z_{t-1}, y_t, y_{t+1}, z_{t+1}, \dots).$$

Apply quasi-stationarity  $t - 1$  times to obtain

$$(z_0, x_t, x_{t+1}, z_{t+1}, \dots) \succ (z_0, y_t, y_{t+1}, z_{t+1}, \dots).$$

By part (b) of initial separability, conclude that

$$(z_0, x_t, x_{t+1}, z'_{t+1}, \dots) \succ (z_0, y_t, y_{t+1}, z'_{t+1}, \dots).$$

By part (c) of initial separability, conclude that

$$(z'_0, x_t, x_{t+1}, z'_{t+1}, \dots) \succ (z'_0, y_t, y_{t+1}, z'_{t+1}, \dots).$$

Apply quasi-stationarity  $t - 1$  times to obtain

$$(z'_0, z'_1, \dots, z'_{t-1}, x_t, x_{t+1}, z'_{t+1}, \dots) \succ (z'_0, z'_1, \dots, z'_{t-1}, y_t, y_{t+1}, z'_{t+1}, \dots).$$

The proof of the independence of  $\{t, t + 1, \dots\}$  for  $t = 2, \dots$  is analogous.

**Step 2.** Show that any period  $t$  is sensitive. To see that, observe that by sensitivity of the period  $t = 1$  there exists  $x \in \mathcal{F}$  and  $c, c' \in C$  such that

$$(x_0, c, x_{t+1}, x_{t+2}, \dots) \succ (x_0, c', x_{t+1}, x_{t+2}, \dots).$$

By quasi-stationarity, applied  $t - 1$  times conclude that

$$(x_0, x_1, \dots, x_{t-1}, c, x_{t+1}, x_{t+2}, \dots) \succ (x_0, x_1, \dots, x_{t-1}, c', x_{t+1}, x_{t+2}, \dots).$$

**Step 3.** Additive representation on  $X_T$ . Fix  $T \geq 1$  and fix  $e \in C$ . Weak Order, Finite Continuity and Steps 1 and 2 imply that (By Theorem 1 of Gorman (1968), together with Vind (1971)) the restriction of  $\succ$  to  $X_T$  is represented by

$$(x_0, x_1, \dots, x_T, c, c, \dots) \mapsto \sum_{t=0}^T v_{t,T}(x_t) + R_T(c)$$

for some nonconstant and continuous maps  $v_{t,T}$  and  $R_T$  from  $C$  to  $\mathbb{R}$ . By the uniqueness of additive representations, the above functions can be chosen to satisfy

$$v_{t,T}(e) = R_T(e) = 0 \tag{A.1}$$

**Step 4.** Since any  $X_T \subseteq X_{T+1}$ , there are two additive representations of  $\succ$  on  $X_T$ :

$$(x_0, x_1, \dots, x_T, c, c, \dots) \mapsto \sum_{t=0}^T v_{t,T}(x_t) + R_T(c)$$

and

$$(x_0, x_1, \dots, x_T, c, c, \dots) \mapsto \sum_{t=0}^T v_{t,T+1}(x_t) + v_{T+1,T+1}(x_t) + R_{T+1}(c).$$

By the uniqueness of additive representations and the normalization (A.1), the above functions must satisfy  $v_{t,T+1}(c) = \gamma_{T+1} v_{t,T}(c)$  for  $t = 0, 1, \dots, T-1$  and  $v_{T+1,T+1}(c) + R_{T+1}(c) = \gamma_{T+1} R_T(c)$  for some  $\gamma_{T+1} > 0$ . By the uniqueness of additive representations the representations can be normalized so that  $\gamma_{T+1} = 1$ . Let  $v_t$  denote the common function  $v_{t,T}$ . With this notation, we obtain

$$v_{T+1}(c) + R_{T+1}(c) = R_T(c). \quad (\text{A.2})$$

**Step 5.** By quasi-stationarity, for any  $T \geq 1$  the two additive representations of  $\succsim$  on  $X_T$ :

$$(e, x_0, x_1, \dots, x_{T-1}, c, c, \dots) \mapsto v_0(e) + \sum_{t=1}^T v_t(x_{t-1}) + R_T(c)$$

and

$$(e, x_0, x_1, \dots, x_{T-1}, c, c, \dots) \mapsto v_0(e) + \sum_{t=1}^T v_{t+1}(x_{t-1}) + R_{T+1}(c)$$

represent the same preference. By the uniqueness of additive representations, and the normalization (A.1), there exists  $\delta_T > 0$  such that for all  $t = 1, 2, \dots$ ,  $v_{t+1}(c) = \delta_T v_t(c)$  for all  $c \in C$  and  $R_{T+1}(c) = \delta_T R_T(c)$ . Note, that  $\delta_T$  is independent of  $T$ , since the functions  $v$  and  $R$  are independent of  $T$ ; let  $\delta$  denote this common value.

**Step 6.** Define  $u := v_0$ ,  $v := \delta^{-1} v_1$  and  $R := \delta^{-2} R_1$ . With this notation, equation (A.2) is  $\delta^{T+1} v(c) + \delta^{T+2} R(c) = \delta^{T+1} R(c)$  for all  $c \in C$ . Observe, that  $\delta = 1$  implies that  $v$  is a constant function, which is a contradiction; hence,  $\delta \neq 1$ . Thus,  $R(c) = \frac{1}{1-\delta} v(c)$  for all  $c \in C$ . Thus, the preference on  $X_T$  is represented by

$$(x_0, x_1, \dots, x_T, c, c, \dots) \mapsto u(x_0) + \sum_{t=1}^T \delta^t v(x_t) + \frac{\delta^{T+1}}{1-\delta} v(c).$$

To rule out  $\delta > 1$  note that since  $v$  is nonconstant, there exist  $a, b \in C$  such that  $v(a) > v(b)$ . Then, since  $\delta + \frac{\delta^2}{1-\delta} < 0$  it follows that  $u(a) + \delta v(b) + \frac{\delta^2}{1-\delta} v(b) > u(a) + \delta v(a) + \frac{\delta^2}{1-\delta} v(a)$ , so  $eb \succ a$ . However, by tail continuity there exists  $T$  such that  $(eb)_{Ta} \succ a$ , which implies that

$$u(a) + (\delta + \dots + \delta^T)v(b) + \frac{\delta^{T+1}}{1-\delta}v(a) > u(a) + (\delta + \dots + \delta^T)v(a) + \frac{\delta^{T+1}}{1-\delta}v(a).$$

Thus,  $(\delta + \dots + \delta^T)(v(b) - v(a)) > 0$  which contradicts  $v(a) > v(b)$  and  $\delta > 0$ . Thus,  $\delta < 1$  and  $U(x)$  represents  $\succ$  on  $X_T$  for any  $T$ .

**Step 7.** Fix  $x \in \mathcal{F}$ . By constant-equivalence, there exists  $c \in C$  with  $x \sim c$ . Suppose there exists  $a \in C$  such that  $c \succ a$ . Then by tail continuity there exists  $\tau$  such that for all  $T \geq \tau$ ,  $x_T a \succ a$ , which by Step 6 implies that  $U(x_T a) > U(a)$ . This implies that

$$\begin{aligned} \exists_\tau \forall_{T \geq \tau} u(x_0) + \sum_{t=1}^T \delta^t v(x_t) + \frac{\delta^{T+1}}{1-\delta} v(a) &> u(a) + \sum_{t=1}^T \delta^t v(a) + \frac{\delta^{T+1}}{1-\delta} v(a) \\ \exists_\tau \forall_{T \geq \tau} \sum_{t=1}^T [\delta^t v(x_t) - \delta^t v(a)] &> [u(a) - u(x_0)] \\ \exists_\tau \inf_{T \geq \tau} \sum_{t=1}^T [\delta^t v(x_t) - \delta^t v(a)] &\geq [u(a) - u(x_0)] \\ \sup_\tau \inf_{T \geq \tau} \sum_{t=1}^T [\delta^t v(x_t) - \delta^t v(a)] &\geq [u(a) - u(x_0)], \end{aligned}$$

which means that  $\liminf_T \sum_{t=1}^T [\delta^t v(x_t) - \delta^t v(a)] \geq [u(a) - u(x_0)]$ . Since the sequence  $\sum_{t=1}^T \delta^t v(a)$  converges, it follows that

$$u(x_0) + \liminf_T \sum_{t=1}^T \delta^t v(x_t) \geq u(a) + \lim_T \sum_{t=1}^T \delta^t v(a) = U(a).$$

Since this is true for all  $a \prec c$ , by connectedness of  $C$  and continuity of  $u$  and  $v$  it follows that

$$u(x_0) + \liminf_T \sum_{t=1}^T \delta^t v(x_t) \geq U(c). \tag{A.3}$$

On the other hand, suppose that  $a \succsim c$  for all  $a \in C$ . Then, by constant-equivalence for all  $T$  there exists  $b \in C$  such that  $x_T c \sim b$ . This implies that  $x_T c \succsim c$ . Thus,

$$\begin{aligned} \forall_T u(x_0) + \sum_{t=1}^T \delta^t v(x_t) + \frac{\delta^{T+1}}{1-\delta} v(c) &\geq u(c) + \sum_{t=1}^T \delta^t v(c) + \frac{\delta^{T+1}}{1-\delta} v(c) \\ \forall_T \sum_{t=1}^T \delta^t v(x_t) - \sum_{t=1}^T \delta^t v(c) &\geq u(c) - u(x_0) \\ \liminf_T \sum_{t=1}^T \delta^t v(x_t) - \sum_{t=1}^T \delta^t v(c) &\geq u(c) - u(x_0) \end{aligned}$$

Since the sequence  $\sum_{t=1}^T \delta^t v(c)$  converges, equation (A.3) follows.

An analogous argument implies that  $\limsup_T \sum_{t=0}^T \delta^t v(x_t) \leq U(c)$ , which establishes the existence of the limit of the partial sums and the representation.  $\square$

### A.1.2 Proof of Theorem 2

We have

$$(e, b, a, \dots) \succsim (e, a, b, \dots)$$

iff

$$u(e) + \delta v(b) + \frac{\delta^2}{1-\delta} v(a) \geq u(e) + \delta v(a) + \frac{\delta^2}{1-\delta} v(b)$$

iff

$$v(b) + \frac{\delta}{1-\delta} v(a) \geq v(a) + \frac{\delta}{1-\delta} v(b)$$

iff

$$[v(b) - v(a)] \frac{1-2\delta}{1-\delta} \geq 0$$

iff

$$1 - 2\delta \leq 0 \quad \square$$

### A.1.3 Proof of Theorem 3

The following lemma is key in the proof of Theorem 3.

**Lemma 3.** *For any  $\delta \in [0.5, 1]$  and any  $\beta \in (0, 1]$  there exists a sequence  $\{\alpha_t\}_t$  of elements in  $\{0, 1\}$  such that  $\beta = \sum_{t=0}^{\infty} \alpha_t \delta^t$ .*

*Proof.* Let  $d_0 := 0$  and  $\alpha_0 := 0$  and define the sequences  $\{d_t\}$  and  $\{\alpha_t\}$  by

$$d_{t+1} := \begin{cases} d_t + \delta^{t+1} & \text{if } d_t + \delta^{t+1} \leq \beta \\ d_t & \text{otherwise.} \end{cases}$$

and

$$\alpha_{t+1} := \begin{cases} 1 & \text{if } d_t + \delta^{t+1} \leq \beta \\ 0 & \text{otherwise.} \end{cases}$$

Since the sequence  $\{d_n\}$  is increasing and bounded from above by  $\beta$ , it must converge; let  $d := \lim d_t$ . It follows that  $d = \sum_{t=0}^{\infty} \alpha_t \delta^t$ . Suppose that  $d < \beta$ . It follows that  $\alpha_t = 1$  for almost all  $t$ ; since otherwise there would exist arbitrarily large  $t$  with  $\alpha_t = 0$ , and since  $\delta^t < \beta - d$  for some such  $t$  that would contradict the construction of the sequence  $\{d_t\}$ . Let  $T := \max\{t : \alpha_t = 0\}$ . We have  $d = d_{T-1} + \frac{\delta^{T+1}}{1-\delta} \leq \beta$ . Since  $\delta \geq 0.5$ , it follows that  $\delta^T \leq \frac{\delta^{T+1}}{1-\delta}$ , so  $d_{T-1} + \delta^T \leq \beta$ , which contradicts the construction of the sequence  $\{d_t\}$ .  $\square$

### Proof of Theorem 3

The necessity of Axioms 1–9 follows from Theorems 1 and 2 and Lemma 3. Suppose that Axioms 1–9 hold. By Theorems 1 and 2 the preference is represented by (1.3) with  $\delta \geq 0.5$ . Normalize  $u$  and  $v$  so that there exists  $\hat{e} \in C$  with  $u(\hat{e}) = v(\hat{e}) = 0$ . Let

$M$  be as in Axiom 9. Define  $\gamma := \sum_{t \in M} \delta^{t-1}$ . Axiom 9 implies that for all  $a, b, c, d \in C$

$$u(a) + \delta v(b) > u(c) + \delta v(d)$$

if and only if

$$v(a) + \gamma v(b) > v(c) + \gamma v(d).$$

By the uniqueness of the additive representations, there exists  $\beta > 0$  and  $\lambda_1, \lambda_2 \in \mathbb{R}$  such that  $v(e) = \beta u(e) + \lambda_1$  and  $\gamma v(e) = \beta \delta v(e) + \lambda_2$  for all  $e \in C$ . By the above normalization,  $\lambda_1 = \lambda_2 = 0$ . Hence,  $v(e) = \beta u(e)$  for all  $e \in C$  and  $\beta = \sum_{t \in M} \delta^{t-2}$ .  $\square$

#### A.1.4 Proof of Theorem 4

The necessity of Axioms 1-7 and 10 is straightforward. For Axiom 11, if  $(b, e_2) \succsim (a, e_1)$ ,  $(c, e_1) \succsim (d, e_2)$  and  $(e_3, a) \sim (e_4, b)$ , it follows that:

$$u(b) + \frac{\delta}{1-\delta} \beta u(e_2) \geq u(a) + \frac{\delta}{1-\delta} \beta u(e_1) \quad (\text{A.4})$$

$$u(c) + \frac{\delta}{1-\delta} \beta u(e_1) \geq u(d) + \frac{\delta}{1-\delta} \beta u(e_2) \quad (\text{A.5})$$

$$u(e_3) + \frac{\delta}{1-\delta} \beta u(a) = u(e_4) + \frac{\delta}{1-\delta} \beta u(b) \quad (\text{A.6})$$

Equations A.4 – A.5 imply  $u(b) - u(a) \geq u(d) - u(c)$ . Suppose that the implication of Axiom 11 does not hold, so that  $(e_4, d) \succ (e_3, c)$ . Then

$$u(e_4) + \frac{\delta}{1-\delta} \beta u(d) > u(e_3) + \frac{\delta}{1-\delta} \beta u(c) \quad (\text{A.7})$$

Since  $0 < \beta, 0 < \delta < 1$ , equations A.6 – A.7 imply  $u(d) - u(c) > u(b) - u(a)$ . A contradiction. By analogy, the second condition of Axiom 11 is also necessary. Therefore, Axiom 11 is satisfied by the representation in Theorem 4.

Now, we prove sufficiency. From Theorem 1 it follows that  $\succsim$  admits the represen-

tation in (1.3). Define the binary relation  $\succsim^*$  over the elements of  $C^2$  as follows:

$$(b, c) \succsim^* (a, d)$$

$\iff$  there exists  $e_1, e_2, e_3, e_4 \in C$  such that

$$(b, e_2) \succsim (a, e_1) \text{ and } (c, e_1) \succsim (d, e_2) \text{ and } (e_3, a) \sim (e_4, b) \quad (\text{A.8})$$

We break the proof of sufficiency into four steps:

**Step 1:** First, we argue that  $\succsim^*$  admits the following additive representation:

$$(b, c) \succsim^* (a, d) \iff u(b) + u(c) \geq u(a) + u(d)$$

Using the definition of  $\succsim^*$  and the representation (1.3) of  $\succsim$ , it follows that  $(b, c) \succsim^* (a, d)$  implies the existence of elements  $e_1, e_2 \in C$  such that:

$$u(a) + \frac{\delta}{1-\delta}v(e_1) \leq u(b) + \frac{\delta}{1-\delta}v(e_2)$$

and

$$u(d) + \frac{\delta}{1-\delta}v(e_2) \leq u(c) + \frac{\delta}{1-\delta}v(e_1)$$

Therefore  $u(b) + u(c) \geq u(a) + u(d)$ .

Now, suppose  $u(b) + u(c) \geq u(a) + u(d)$ . We consider the following 6 cases and we show that Condition A.8 is satisfied.

1.  $u(b) \geq u(a), u(c) \geq u(d), v(a) \geq v(b)$ : Set  $e = e_1 = e_2$  for any  $e \in C$ , and choose  $e_3, e_4$  to satisfy  $u(e_3) + \frac{\delta}{1-\delta}v(a) = u(e_4) + \frac{\delta}{1-\delta}v(b)$ . Then, Condition (A.8) is satisfied.
2.  $u(b) \geq u(a), u(c) \geq u(d), v(a) < v(b)$ : Set  $e = e_1 = e_2$  for any  $e \in C$  and choose  $e_3, e_4$  to have  $u(e_3) + \frac{\delta}{1-\delta}v(a) = u(e_4) + \frac{\delta}{1-\delta}v(b)$ . Again, condition A.8 is satisfied



and  $(b, c) \succsim^* (a, d)$ .

3.  $u(b) \geq u(a), u(c) < u(d), v(a) \geq v(b)$ : Note that  $u(b) - u(a) \geq u(d) - u(c) > 0$ .

Find  $e_1, e_2$  to satisfy:  $\frac{\delta}{1-\delta}[v(e_1) - v(e_2)] = u(d) - u(c) > 0$ . And set  $e = e_3, e_4$  to get indifference.

4.  $u(b) \geq u(a), u(c) < u(d), v(a) < v(b)$ : Do the same as above.

5.  $u(b) < u(a), u(c) \geq u(d), v(a) \geq v(b)$ : Find  $e_1, e_2$  to satisfy:  $\frac{\delta}{1-\delta}[v(e_1) - v(e_2)] = u(b) - u(a) < 0$ . Note that

$$0 = u(b) - u(a) - \frac{\delta}{1-\delta}[v(e_1) - v(e_2)] \geq u(d) - u(c) - \frac{\delta}{1-\delta}[v(e_1) - v(e_2)]$$

6.  $u(b) < u(a), u(c) \geq u(d), v(a) < v(b)$ : Do the same as above.

In any event  $u(b) + u(c) \geq u(a) + u(d)$  implies  $(b, c) \succsim^* (a, d)$ . Therefore, the preference relation  $\succsim^*$  admits an additive representation in terms of  $u$ .

**Step 2:** The preference relation  $\succsim^*$  also admits a representation in terms of the index  $v$ :

$$(b, c) \succsim^* (a, d) \iff v(b) + v(c) \geq v(a) + v(d)$$

Using the definition of  $\succsim^*$  and Axiom 11 it follows that:

$$u(e_3) + \frac{\delta}{1-\delta}v(a) = u(e_4) + \frac{\delta}{1-\delta}v(b)$$

and

$$u(e_3) + \frac{\delta}{1-\delta}v(c) \geq u(e_4) + \frac{\delta}{1-\delta}v(d)$$

which implies  $v(b) + v(c) \geq v(a) + v(d)$ . Now, for the other direction, we proceed as in Step 1. Suppose  $v(b) + v(c) \geq v(a) + v(d)$ . Proceeding exactly as before, there are elements  $e_1, e_2, e_3, e_4$  such that  $(e_2, b) \succsim (e_1, a)$ ,  $(e_1, c) \succsim (e_2, d)$  and  $(a, e_3) \sim (b, e_4)$ .

By Axiom 11, it follows that  $(c, e_3) \succsim (b, e_4)$ . And therefore,  $u(b) + u(c) \geq u(a) + u(d)$ . Therefore,  $(b, c) \succsim^* (a, d) \iff v(b) + v(c) \geq v(a) + v(d)$ .

**Step 3:** Since the preference relation  $\succsim^*$  admits two different additive representations it follows that the two utility indexes are related through a monotone affine transformation. This is, there exists  $\beta > 0$  and  $\gamma$  such that for all  $a \in C$ :

$$v(a) = \beta u(a) + \gamma$$

We conclude that  $\succsim$  is represented by the mapping

$$x \mapsto u(x_0) + \beta \sum_{t=1}^{\infty} \delta^t u(x_t). \quad (\text{A.9})$$

with  $\beta > 0$ .

**Step 4:** Take  $a, c \in C$  such that  $u(a) > u(c)$ . The existence of such an element follows from the sensitivity axiom. Choose  $b, d$  to satisfy:

$$u(a) + \delta u(b) = u(c) + \delta u(d)$$

Axiom 10 implies that

$$u(a) + \beta \delta u(b) \geq u(c) + \beta \delta u(d)$$

The two inequalities imply  $\beta \leq 1$ .

### A.1.5 Proof of Theorem 5

**Remark 16.** Both Ghirardato and Marinacci (2001) and Nakamura (1990) study Choquet preferences, so their axioms have comonotonicity requirements. To have simpler statements and to avoid introducing the concept of comonotonicity in the main text

we use stronger axioms that hold for all, not necessarily comonotone acts, but the comonotone versions of those axioms could be used (are equivalent in the presence of other axioms).

### Proof of Theorem 5

The necessity of the axioms is straightforward. For sufficiency, we rely on the work of Ghirardato and Marinacci (2001). Note that their axiom B1 follows from our axioms 1 and 2. Their axioms B2 and B3 follow from our axiom 13. Their axiom S1 follows from the fact that by Theorem 1 the functions  $u$  and  $v$  are continuous. Finally their axiom S2 follows from our axiom 12. Thus, by their Lemma 31 there exists  $\alpha \in (0, 1)$  and  $w : C \rightarrow \mathbb{R}$  such that  $(a, b) \mapsto \alpha w(a) + (1 - \alpha)w(b)$  represents  $\succsim$ . By uniqueness of additive representations,  $w$  is a positive affine transformation of  $u$ . Step 4 in the proof of Theorem 5 concludes the proof.  $\square$

### Nakamura's axiom

An alternative to Theorem 5 is the following:

**Axiom 14.** (Nakamura's A6) For  $a, b, c, d \in C$  such that  $b \succsim a$ ,  $d \succsim c$ ,  $d \succsim b$  and  $c \succsim a$ :

$$(c(a, b), c(c, d)) \sim (c(a, c), c(b, d))$$

and

$$(c(c, d), c(a, b)) \sim (c(c, a), c(d, b))$$

**Theorem 1.** *The preference  $\succsim$  satisfies Axioms 1–7 and 13–14 if and only if there exists a nonconstant and continuous function  $u : C \rightarrow \mathbb{R}$  and parameters  $\beta > 0$  and*

$\delta \in (0, 1)$  such that  $\succsim$  is represented by the mapping

$$x \mapsto u(x_0) + \beta \sum_{t=1}^{\infty} \delta^t u(x_t).$$

Moreover, it satisfies Axiom 10 if and only if  $\beta \leq 1$ , i.e.,  $\succsim$  has the quasi-hyperbolic discounting representation.

*Proof.* The necessity of Axioms 1-7, 10 and 13 is straightforward. For Axiom 14, take  $a, b, c, d \in C$  as in the statement of the axiom and note that:

$$c(a, b) \equiv c_1, \quad u(c_1) + \frac{\delta}{1-\delta} \beta u(c_1) = u(a) + \frac{\delta}{1-\delta} \beta u(b) \quad (\text{A.10a})$$

$$c(c, d) \equiv c_2, \quad u(c_2) + \frac{\delta}{1-\delta} \beta u(c_2) = u(c) + \frac{\delta}{1-\delta} \beta u(d) \quad (\text{A.10b})$$

And also,

$$c(a, c) \equiv c_3, \quad u(c_3) + \frac{\delta}{1-\delta} \beta u(c_3) = u(a) + \frac{\delta}{1-\delta} \beta u(c) \quad (\text{A.11a})$$

$$c(b, d) \equiv c_4, \quad u(c_4) + \frac{\delta}{1-\delta} \beta u(c_4) = u(b) + \frac{\delta}{1-\delta} \beta u(d) \quad (\text{A.11b})$$

Therefore, using equations A.10a–b

$$[1 + \beta \frac{\delta}{1-\delta}] [u(c_1) + \frac{\delta}{1-\delta} u(c_2)] = u(a) + \frac{\delta}{1-\delta} \beta u(b) + \frac{\delta}{1-\delta} \beta u(c) + \left( \frac{\delta}{1-\delta} \beta \right)^2 u(d) \quad (\text{A.12})$$

and using A.11a–b

$$[1 + \beta \frac{\delta}{1-\delta}] [u(c_3) + \frac{\delta}{1-\delta} u(c_4)] = u(a) + \frac{\delta}{1-\delta} \beta u(c) + \frac{\delta}{1-\delta} \beta u(b) + \left( \frac{\delta}{1-\delta} \beta \right)^2 u(d) \quad (\text{A.13})$$

So,  $(c_1, c_2) \sim (c_3, c_4)$ . The second implication of Axiom 14 follows by analogy.

For sufficiency of the axioms we rely on the proof of Lemma 3 (Proposition 1) in

Nakamura (1990)'s.<sup>1</sup> The argument goes as follows. Consider the restriction of  $\succsim$  to elements of the form  $(a, b)$ , with  $a, b \in C$  and  $b \succsim a$ . Denote it by  $\succsim_R$ . The proof of Theorem 1 implies Lemma 2 (Part 1 and 2) of Nakamura (1990), with  $S = (s_1, s_2)$ ,  $A = s_1$ ,  $\phi \equiv u$  and  $\psi \equiv \frac{\delta}{1-\delta}v$ . Our axioms 13 and 14 coincide exactly with A3 and A6 in Nakamura (1990) when  $S = (s_1, s_2)$ . Therefore, Lemma 3 implies there is a real valued function  $r(x)$  such that:

$$(a, b) \succsim_R (c, d) \iff \alpha r(a) + (1 - \alpha)r(b) \geq \alpha r(c) + (1 - \alpha)r(d)$$

where  $r$  is defined (pg. 356 Nakamura (1990)) as  $\phi(c)/\alpha$  for all  $c \in C$  and  $\alpha = 1/(1+\beta^*)$ , with  $\beta^*$  such that  $\psi(c) = \beta^*\phi(c) + \gamma^*$ ,  $\beta^* > 0$ . Hence, it follows that for every  $c \in C$ ,  $\frac{\delta}{1-\delta}v(c) = \beta^*u(c) + \gamma^*$ . If we set  $\beta = \frac{1}{\beta^*}$ , then we get  $u(c) = \frac{\delta}{1-\delta}\beta u(c) + \gamma$ . The representation (1.3) becomes:

$$x \mapsto u(x_0) + \beta \sum_{t=1}^{\infty} \delta^t u(x_t), \quad \beta > 0.$$

Step 4 in the proof of Theorem 5 concludes the proof. □

---

<sup>1</sup>Nakamura's results are used explicitly by Chew and Karni (1994) and implicitly by Ghirardato and Marinacci (2001).

## A.2 Proofs for Chapter 2

### A.2.1 Proof of Lemma 1

First note the preliminary result that under Assumptions  $L_\Pi$  and HL

$$\frac{1}{\sqrt{S}} \begin{pmatrix} \mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'\mathbf{Y} \end{pmatrix} = \begin{pmatrix} \beta\mathbf{C} + \mathbf{Z}'\mathbf{v}_1/\sqrt{S} \\ \mathbf{C} + \mathbf{Z}'\mathbf{v}_2/\sqrt{S} \end{pmatrix} \quad (\text{A.14})$$

$$\xrightarrow{d} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \quad (\text{A.15})$$

1.  $\hat{\beta}_{TSLs} \equiv (\mathbf{Y}'\mathbf{P}_Z\mathbf{Y})^{-1}(\mathbf{Y}'\mathbf{P}_Z\mathbf{y}) = (\mathbf{Y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y})^{-1}(\mathbf{Y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y})$ . Since we have assumed that  $\mathbf{Z}'\mathbf{Z}/S = \mathbb{I}_K$ , the result follows from (A.15) and the continuous mapping theorem.

2. Write  $\mathbf{J} = \begin{bmatrix} 1 & 0 \\ -\beta & 1 \end{bmatrix}$  and  $\kappa = S(k-1)$ . Note that  $\mathbf{J}$  is nonsingular and so the roots of  $|\mathbf{J}'[\mathbf{y}, \mathbf{Y}]'[\mathbf{y}, \mathbf{Y}] - k[\mathbf{y}, \mathbf{Y}]'\mathbf{M}_Z[\mathbf{y}, \mathbf{Y}]| = 0$  are the same as of

$$|\mathbf{J}'[\mathbf{y}, \mathbf{Y}]'[\mathbf{y}, \mathbf{Y}]\mathbf{J} - k\mathbf{J}'[\mathbf{y}, \mathbf{Y}]'\mathbf{M}_Z[\mathbf{y}, \mathbf{Y}]\mathbf{J}| = 0. \text{ Moreover}$$

$$[\mathbf{y}, \mathbf{Y}]'[\mathbf{y}, \mathbf{Y}] - (1 + \kappa/S)[\mathbf{y}, \mathbf{Y}]'\mathbf{M}_Z[\mathbf{y}, \mathbf{Y}] = [\mathbf{y}, \mathbf{Y}]'\mathbf{P}_Z[\mathbf{y}, \mathbf{Y}] - \kappa[\mathbf{y}, \mathbf{Y}]'\mathbf{M}_Z[\mathbf{y}, \mathbf{Y}]/S$$

$$\xrightarrow{d} [\gamma_1, \gamma_2]'[\gamma_1, \gamma_2] - \kappa\boldsymbol{\Omega} \text{ uniformly in } \kappa \text{ over compact sets. The solutions of}$$

$$|[\mathbf{y}, \mathbf{Y}]'[\mathbf{y}, \mathbf{Y}] - (1 + \kappa/S)[\mathbf{y}, \mathbf{Y}]'\mathbf{M}_Z[\mathbf{y}, \mathbf{Y}]| = 0 \text{ therefore converge to those of}$$

$$|\mathbf{J}'[\gamma_1, \gamma_2]'[\gamma_1, \gamma_2]\mathbf{J} - \kappa\mathbf{J}'\boldsymbol{\Omega}\mathbf{J}| = 0. \text{ With } \mathbf{J}'\boldsymbol{\Omega}\mathbf{J} = \boldsymbol{\Sigma} \text{ thus } S(\hat{k}_{LIML} - 1) \xrightarrow{d} \kappa_{LIML}$$

where  $\kappa_{LIML}$  is as given in Lemma 1.2.

Then  $\hat{\beta}_{LIML} - \beta =$

$$\begin{aligned} & \left[ \mathbf{Y}'(\mathbb{I}_S - \hat{k}_{LIML}\mathbf{M}_Z)\mathbf{Y} \right]^{-1} [\mathbf{Y}'(\mathbb{I}_S - \hat{k}_{LIML}\mathbf{M}_Z)(\mathbf{y} - \beta\mathbf{Y})] \\ &= \left[ \mathbf{Y}'\mathbf{P}_Z\mathbf{Y} - S(\hat{k}_{LIML} - 1)\frac{\mathbf{Y}'\mathbf{M}_Z\mathbf{Y}}{S} \right]^{-1} [\mathbf{Y}'\mathbf{P}_Z(\mathbf{y} - \beta\mathbf{Y}) - S(\hat{k}_{LIML} - 1)\frac{\mathbf{Y}'\mathbf{M}_Z(\mathbf{y} - \beta\mathbf{Y})}{S}] \\ & \xrightarrow{d} [\gamma_2'\gamma_2 - \kappa_{LIML}\sigma_2^2]^{-1} [\gamma_2(\gamma_1 - \beta\gamma_2) - \kappa_{LIML}\sigma_{12}] \end{aligned}$$

3. Note that  $\hat{\omega}_2^2 \equiv (\mathbf{Y} - \mathbf{P}_Z \mathbf{Y})'(\mathbf{Y} - \mathbf{P}_Z \mathbf{Y})/(S - K - 1) = (\mathbf{v}_2 - \mathbf{P}_Z \mathbf{v}_2)'(\mathbf{v}_2 - \mathbf{P}_Z \mathbf{v}_2)/(S - K - 1) \xrightarrow{d} \omega_2^2$  by Assumptions  $L_\Pi$  and HL. The result follows from (A.15) and the continuous mapping theorem.
4. and 5. follow from (A.15), the continuous mapping theorem, and Assumptions  $L_\Pi$  and HL.

### A.2.2 LIML Distribution in Illustrative Example

We show that in the illustrative example heteroskedascity and serial correlation can effectively make instruments weaker for LIML. Assume  $\mathbf{W} = a^2 \boldsymbol{\Omega} \otimes \mathbb{I}_K$ . Remember that  $\hat{\beta}_{LIML} = \arg \min_{\tilde{\beta}} (\mathbf{y} - \tilde{\beta} \mathbf{Y})' \mathbf{P}_Z (\mathbf{y} - \tilde{\beta} \mathbf{Y}) / (\mathbf{y} - \tilde{\beta} \mathbf{Y})' (\mathbf{y} - \tilde{\beta} \mathbf{Y})$ . We will analyze the weak instrument limit of the LIML objective function. Note that, using assumptions  $L_\Pi$  and HL  $\mathbf{Z}'(\mathbf{y} - \tilde{\beta} \mathbf{Y})/\sqrt{S} \xrightarrow{d} \boldsymbol{\gamma}_1 - \tilde{\beta} \boldsymbol{\gamma}_2$ .

Moreover,  $(\mathbf{y} - \tilde{\beta} \mathbf{Y})'(\mathbf{y} - \tilde{\beta} \mathbf{Y})/S \xrightarrow{p} \omega_1^2 - 2\tilde{\beta}\omega_{12} + \tilde{\beta}^2\omega_2^2$  uniformly in  $\tilde{\beta}$  over compact sets. Hence  $\beta_{LIML}^*$  is distributed according to

$$\arg \min_{\tilde{\beta}} a^2 \frac{(\omega_1 \boldsymbol{\psi}_1 - \tilde{\beta} \omega_2 \boldsymbol{\psi}_2)' (\omega_1 \boldsymbol{\psi}_1 - \tilde{\beta} \omega_2 \boldsymbol{\psi}_2)}{\omega_1^2 - 2\tilde{\beta}\omega_{12} + \tilde{\beta}^2\omega_2^2}$$

Just as for the  $\beta_{TSLS}$ , the vector of first stage coefficients  $\mathbf{C}$  and the parameter  $a$  enter into the asymptotic distribution  $\beta_{LIML}^*$  only through the noncentrality parameter  $\mathbf{C}'\mathbf{C}/(a^2\omega_2^2)$ .

### A.2.3 Proof of Theorem 1

#### A.3.1 Proof of Theorem 1.1

We follow Rothenberg (1984) in developing the Nagar (1959) moments for the TSLS and LIML estimators. We need to expand  $\beta_{TSLS}^*$  and  $\beta_{LIML}^*$  as second order Taylor expansions in  $\mu^{-1}$  around  $\mu^{-1} = 0$ .

We start by developing the Taylor expansion for  $\kappa_{LIML}$ . Write  $\mathbf{z}_u = \mathbf{S}_1^{-1/2}(\gamma_1 - \beta\gamma_2)$  and  $\mathbf{z}_v = \mathbf{S}_2^{-1/2}(\gamma_2 - C)$  so  $\mathbf{z}_u$  and  $\mathbf{z}_v$  are standard multivariate normal. Also write  $\boldsymbol{\lambda} = \mu \text{tr}(\mathbf{S}_2)^{1/2} \mathbf{S}_2^{-1/2} \mathbf{C}_0$  where  $\mathbf{C}_0 = \mathbf{C} / \|\mathbf{C}\|$ .

$\kappa_{LIML}$  is defined as the smallest root of the determinantal equation

$$\det \left( \mathbf{A} - \kappa_{LIML} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right) = 0 \quad (\text{A.16})$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{z}_u' \mathbf{S}_1 \mathbf{z}_u & \mathbf{z}_u' \mathbf{S}_1^{1/2} \mathbf{S}_2^{1/2} (\boldsymbol{\lambda} + \mathbf{z}_v) \\ \mathbf{z}_u' \mathbf{S}_1^{1/2} \mathbf{S}_2^{1/2} (\boldsymbol{\lambda} + \mathbf{z}_v) & (\mathbf{z}_v + \boldsymbol{\lambda})' \mathbf{S}_2 (\mathbf{z}_v + \boldsymbol{\lambda}) \end{bmatrix}$$

We can rewrite this as a quadratic equation

$$\left( \frac{\kappa_{LIML}}{\mu^2} \right)^2 - \frac{\sigma_v^2 a_{11} + \sigma_u^2 a_{22} - 2a_{12} \sigma_{12}}{\mu^2 \det \boldsymbol{\Sigma}} \frac{\kappa_{LIML}}{\mu^2} + \frac{\det \mathbf{A}}{\mu^4 \det \boldsymbol{\Sigma}} = 0 \quad (\text{A.17})$$

We use the method of undetermined coefficients. Write

$$\kappa_{LIML} \mu^{-2} = c_0 + c_1 \mu^{-1} + c_2 \mu^{-2} + O(\mu^{-3}) \quad (\text{A.18})$$

for unknown constants  $c_0, c_1, c_2$ . Similarly write

$$d(\mu) \equiv \frac{\sigma_v^2 a_{11} + \sigma_u^2 a_{22} - 2a_{12} \sigma_{12}}{\mu^2 \det \boldsymbol{\Sigma}} = d_0 + d_1 \mu^{-1} + d_2 \mu^{-2} + O(\mu^{-3}) \quad (\text{A.19})$$

$$e(\mu) \equiv \frac{\det \mathbf{A}}{\mu^4 \det \boldsymbol{\Sigma}} = \frac{\det \mathbf{A}}{\mu^4 \det \boldsymbol{\Sigma}} = e_0 + e_1 \mu^{-1} + e_2 \mu^{-2} + O(\mu^{-3}) \quad (\text{A.20})$$



where the Taylor series expansions for  $d$  and  $e$  give  $d_0 = \sigma_1^2 \text{tr}(\mathbf{S}_2) / \det \mathbf{\Sigma}$ ,  $e_0 = 0$ ,  $e_1 = 0$ , and  $e_2 = \text{tr}(\mathbf{S}_2) \left[ \mathbf{z}'_u \mathbf{S}_1 \mathbf{z}_u - \left( \mathbf{z}'_u \mathbf{S}_1^{1/2} \mathbf{C}_0 \right)^2 \right] / \det \mathbf{\Sigma}$ .

Substituting (A.18), (A.19) and (A.20) into the quadratic equation (A.17) and equating coefficients gives  $c_0(c_0 - d_0) = 0$ . Since we are interested in the smaller solution, we have  $c_0 = 0$ . Then  $c_0 = 0, c_1 = 0, c_2 = \varepsilon_2/d_0$  and so  $\kappa_{LIML} \mu^{-2} = \frac{1}{\sigma_1^2} \left[ \mathbf{z}'_u \mathbf{S}_1 \mathbf{z}_u - \left( \mathbf{z}'_u \mathbf{S}_1^{1/2} \mathbf{C}_0 \right)^2 \right] \mu^{-2} + O(\mu^{-3})$

We then expand  $\beta_{LIML}^*$

$$\begin{aligned} \beta_{LIML}^* &= \mu^{-1} \frac{\mathbf{C}'_0 \mathbf{S}_1^{1/2} \mathbf{z}_u \text{tr}(\mathbf{S}_2)^{1/2}}{\text{tr}(\mathbf{S}_2)} + \\ &+ \mu^{-2} \frac{1}{\text{tr}(\mathbf{S}_2)} \left( \mathbf{z}'_v \mathbf{S}_2^{1/2} \mathbf{S}_1^{1/2} \mathbf{z}_u - 2 \left( \mathbf{C}'_0 \mathbf{S}_1^{1/2} \mathbf{z}_u \right) \left( \mathbf{C}'_0 \mathbf{S}_2^{1/2} \mathbf{z}_v \right) - c_2 \sigma_{12} \right) \\ &+ O(\mu^{-3}) \end{aligned}$$

Taking the expectation of the first two terms gives the LIML Nagar bias as in the Theorem.

We can similarly derive the Taylor expansion for  $\beta_{TSLs}^*$  according to

$$\begin{aligned} \beta_{TSLs}^* &= \mu^{-1} \frac{\mathbf{C}'_0 \mathbf{S}_1^{1/2} \mathbf{z}_u \text{tr}(\mathbf{S}_2)^{1/2}}{\text{tr}(\mathbf{S}_2)} + \\ &+ \mu^{-2} \frac{1}{\text{tr}(\mathbf{S}_2)} \left( \mathbf{z}'_v \mathbf{S}_2^{1/2} \mathbf{S}_1^{1/2} \mathbf{z}_u - 2 \left( \mathbf{C}'_0 \mathbf{S}_1^{1/2} \mathbf{z}_u \right) \left( \mathbf{C}'_0 \mathbf{S}_2^{1/2} \mathbf{z}_v \right) \right) \\ &+ O(\mu^{-3}) \end{aligned}$$

The Nagar bias is defined as the expected value of the first two terms, and hence equals

$$N_{TSLs}(\beta, \mathbf{C}, \mathbf{W}, \mathbf{\Omega}) = \frac{1}{\text{tr} \mathbf{S}_2} (\text{tr} \mathbf{S}_{12} - 2 \mathbf{C}'_0 \mathbf{S}_{12} \mathbf{C}_0) \mu^{-2}$$

### A.3.2 Proof of Theorems 1.2 and 1.3

We prove Theorem 1.3 first. We assume that  $\mathbf{W}$  and  $\mathbf{\Omega}$  are positive definite, so  $\mathbf{S}$  and  $\mathbf{\Sigma}$  are also positive definite.  $\mathbf{S}_{12}$  is real valued but not necessarily symmetric. Note that

$$tr\mathbf{S}_{12} - 2\mathbf{C}'_0\mathbf{S}_{12}\mathbf{C}_0 = tr\mathbf{S}_{12}^{sym} - 2\mathbf{C}'_0\mathbf{S}_{12}^{sym}\mathbf{C}_0$$

where  $\mathbf{S}_{12}^{sym} = \frac{1}{2}(\mathbf{S}_{12} + \mathbf{S}'_{12})$  is the symmetric part of  $\mathbf{S}_{12}$ . Write  $\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_K \end{bmatrix}$

for the diagonal matrix of eigenvalues of  $\mathbf{S}_{12}^{sym}$ . Assume the eigenvalues are ordered so  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K$ . For any real matrix  $\mathbf{M}$  we write  $|\mathbf{M}| = \sqrt{\mathbf{M}'\mathbf{M}}$  so the Schatten 1-norm for matrices is defined as  $\|\mathbf{M}\|_1 = tr|\mathbf{M}|$ .

$$\begin{aligned} tr\mathbf{S}_{12}^{sym} - 2\mathbf{C}'_0\mathbf{S}_{12}^{sym}\mathbf{C}_0 &\leq \sum_{k=1}^K \lambda_k - 2\lambda_K \\ &= \sum_{k=1}^{K-1} \lambda_k - \lambda_K \\ &\leq \sum_{k=1}^K |\lambda_k| \\ &= \|\mathbf{S}_{12}^{sym}\|_1 \end{aligned}$$

Similarly  $tr\mathbf{S}_{12}^{sym} - 2\mathbf{C}'_0\mathbf{S}_{12}^{sym}\mathbf{C}_0 \geq -\|\mathbf{S}_{12}^{sym}\|_1$ . Hence  $|tr\mathbf{S}_{12}^{sym} - 2\mathbf{C}'_0\mathbf{S}_{12}^{sym}\mathbf{C}_0| \leq \|0.5\mathbf{S}_{12} + 0.5\mathbf{S}'_{12}\|_1 \leq \|\mathbf{S}_{12}\|_1$ . The last step follows from the triangle inequality and from the fact that the eigenvalues of  $\mathbf{S}'_{12}\mathbf{S}_{12}$  and  $\mathbf{S}_{12}\mathbf{S}'_{12}$  are the same.

Now  $tr(\mathbf{S}'_{12}\mathbf{S}_2^{-1}\mathbf{S}_{12}) \leq tr(\mathbf{S}_1)$ , see e.g. Theorem 7.14 in Zhang (2010). By the matrix

trace Cauchy inequality (Liu and Neudecker (1995), Theorem 1) then

$$\begin{aligned}
\|\mathbf{S}_{12}\|_1^2 &= (tr|\mathbf{S}_{12}|)^2 \\
&\leq tr\mathbf{S}_2 tr(|\mathbf{S}_{12}|'\mathbf{S}_2^{-1}|\mathbf{S}_{12}|) \\
&= tr\mathbf{S}_2 tr(\mathbf{S}_{12}'\mathbf{S}_2^{-1}\mathbf{S}_{12})
\end{aligned}$$

Putting this together, we get  $\|\mathbf{S}_{12}\|_1 \leq \sqrt{tr\mathbf{S}_1 tr\mathbf{S}_2}$ , proving Theorem 1.3.

The TSLS part of Theorem 1.2 follows from Theorem 1.3. For the LIML part note that  $B_{LIML}(\mathbf{W}, \mathbf{\Omega}) = \sup_{\beta \in \mathbb{R}} g_{LIML}(\beta)$  where

$$g_{LIML}(\beta) = \max \left( \left| \frac{tr\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2} tr\mathbf{S}_1 - maxeval\mathbf{M}_B}{\sqrt{tr\mathbf{S}_1} \sqrt{tr\mathbf{S}_2}} \right|, \left| \frac{tr\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2} tr\mathbf{S}_1 - mineval\mathbf{M}_B}{\sqrt{tr\mathbf{S}_1} \sqrt{tr\mathbf{S}_2}} \right| \right) \quad (\text{A.21})$$

where  $\mathbf{M}_B = \frac{1}{2}(2\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2}\mathbf{S}_1) + \frac{1}{2}(2\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2}\mathbf{S}_1)'$  and

$$g_{LIML}(\beta) \rightarrow \frac{maxeval\mathbf{W}_2}{tr\mathbf{W}_2} \text{ as } \beta \rightarrow \pm\infty \quad (\text{A.22})$$

For  $\mathbf{W}$  and  $\mathbf{\Omega}$  nonsingular  $g_{LIML}$  is continuous in  $\beta$  everywhere, and hence bounded.

#### A.2.4 Proof of Lemma 2

Assume that  $\mathbf{W}$  and  $\mathbf{\Omega}$  are nonsingular. We prove that the test that rejects if:

$$\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau) \quad (\text{A.23})$$

is asymptotically valid, i.e. its asymptotic size is at most  $\alpha$ .

**Claim 1:** The function  $F_{\mathbf{C}, \mathbf{W}_2}^{-1}(\alpha)$  is continuous in  $\{\mathbf{C}, \mathbf{W}_2\}$ .

**Proof:** :  $\gamma_2'\gamma_2/tr(\mathbf{W}_2)$  is a continuous random variable with nonzero density on  $\mathbb{R}_+$ , and therefore  $F_{\mathbf{C}, \mathbf{W}_2}^{-1}(\alpha)$  is strictly decreasing and continuous in  $\alpha$  everywhere. By Van der Vaart (2000, Lemma 21.2) the quantile function  $F_{\mathbf{C}, \mathbf{W}_2}^{-1}(\alpha)$  is continuous in  $\{\mathbf{C}, \mathbf{W}_2\}$  for any fixed  $\alpha$ .

**Claim 2:** The function  $B_e(\mathbf{W}, \mathbf{\Omega})$  is lower semicontinuous.

**Proof:** The function  $\|n_e(\beta, \mathbf{C}_0, \mathbf{W}, \mathbf{\Omega})\| / BM(\beta, \mathbf{W})$  is continuous in  $\mathbf{W}$  and  $\mathbf{\Omega}$ .  $B_e(\mathbf{W}, \mathbf{\Omega})$  is the supremum of continuous functions, and therefore is lower semicontinuous (Yeh, 2000, p. 274).

**Claim 3:** The function  $c(\alpha, \mathbf{W}_2, x)$  is lower semicontinuous in  $\{\mathbf{W}_2, x\}$ .

**Proof:** The function  $\mathbb{1}_{\mathbf{C}'\mathbf{C}/tr(\mathbf{W}_2) < x}$  is an indicator function of an open set, and therefore lower semicontinuous in  $\{\mathbf{W}_2, x\}$ . The function  $F_{\mathbf{C}, \mathbf{W}_2}^{-1}(\alpha)$  is continuous in  $\mathbf{W}_2$  and greater than 0. Hence the product  $F_{\mathbf{C}, \mathbf{W}_2}^{-1}(\alpha) \mathbb{1}_{\mathbf{C}'\mathbf{C}/tr(\mathbf{W}_2) < x}$  is lower semicontinuous in  $(\mathbf{W}_2, x)$  for any fixed  $\alpha$ .  $c(\alpha, \mathbf{W}_2, x)$  is a supremum of lower semicontinuous functions, and therefore lower semicontinuous in  $(\mathbf{W}_2, x)$  (Yeh, 2000, p. 274).  $c(\alpha, \mathbf{W}_2, x)$  is also clearly nondecreasing in  $x$ .

**Proof of Result:** From the lower semicontinuity of  $B(\mathbf{W}, \mathbf{\Omega})$  and the continuous mapping theorem, it follows that  $\min(B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}}), B_e(\mathbf{W}, \mathbf{\Omega})) \xrightarrow{p} B_e(\mathbf{W}, \mathbf{\Omega})$ . Similarly, for any  $(\widehat{\mathbf{W}}_2, \widehat{x}) \xrightarrow{p} (\mathbf{W}_2, x)$ , the continuous mapping theorem implies that  $\min(c(\alpha, \widehat{\mathbf{W}}_2, \widehat{x}), c(\alpha, \mathbf{W}_2, x)) \xrightarrow{p} c(\alpha, \mathbf{W}_2, x)$ . Then

$$\mathbb{P}\left(\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau)\right) \quad (\text{A.24})$$

$$\leq \mathbb{P}\left(\widehat{F}_{eff} > \min\left(c\left(\alpha, \mathbf{W}_2, \frac{B_e(\mathbf{W}, \mathbf{\Omega})}{\tau}\right), \right. \quad (\text{A.25})$$

$$\left. c\left(\alpha, \widehat{\mathbf{W}}_2, \frac{\min(B_e(\mathbf{W}, \mathbf{\Omega}), B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}}))}{\tau}\right)\right) \quad (\text{A.26})$$

$$\rightarrow \mathbb{P}\left(\widehat{F}_{eff}^* > c\left(\alpha, \mathbf{W}_2, \frac{B_e(\mathbf{W}, \mathbf{\Omega})}{\tau}\right)\right) \quad (\text{A.27})$$

$$= \alpha \quad (\text{A.28})$$

Now we prove the second part of the Lemma. We first prove a bound for the critical values. Let  $F_{\chi_d^2(x)}^{-1}(\alpha)$  the upper  $\alpha$  point of a non-central  $\chi^2$  with  $d$  degrees of freedom and noncentrality parameter  $x$ . For any  $\alpha \in [0, 1]$

$$c(\alpha, \mathbf{W}_2, x) \leq x^* \equiv \left( \sqrt{\max \left( F_{\chi_1^2(0)}^{-1}(\alpha), F_{\chi_2^2(0)/2}^{-1}(\alpha), \dots, F_{\chi_K^2(0)/K}^{-1}(\alpha) \right)} + \sqrt{x} \right)^2.$$

Let  $X_i \sim N(0, 1)$  i.i.d.,  $i = 1, 2, \dots, K$ , and let  $c \in A$  where  $A = \{c \in \mathbb{R}^K \mid \sum_{i=1}^K c_i = 1, \quad c_i \geq 0, \quad \forall i\}$ . From Szekely and Bakirov (2003)  $\tilde{x} \in \mathbb{R}_+$  that  $\inf_{c \in A} P(\sum_{i=1}^K c_i X_i^2 \leq \tilde{x}) = P(\chi_n^2/n(\tilde{x}) \leq \tilde{x})$ , where the function  $n(\tilde{x})$  is integer, non-decreasing, bounded by  $K$  and equal to 1 whenever  $\tilde{x} > 1.536$ . Let  $Q = \sum_{i=1}^K c_i (X_i + b_i)^2$  a quadratic form in normal random variables and write  $\sum_{i=1}^K c_i b_i^2 = \mu^2$ . From the triangle inequality

$$\mathbb{P}[Q > x] = \mathbb{P}\left[\sum_{i=1}^K c_i (X_i + b_i)^2 > x\right] \leq \mathbb{P}\left[\left(\sqrt{\sum_{i=1}^K c_i X_i^2} + \mu\right)^2 > x\right]$$

Whenever  $x > \mu^2$  then  $\mathbb{P}[Q > x] \leq \mathbb{P}\left[\chi_{n(x_1)}^2/n(x_1) > x_1(\mu^2, x)\right]$ , where  $x_1(\mu^2, x) = (x^{1/2} - \mu)^2$ . Moreover, this bound is increasing in  $\mu^2$  whenever  $x > \mu^2$ . Let  $x^*$  as above.

Then  $x_1(x, x^*) = \max \left( F_{\chi_1^2(0)}^{-1}(\alpha), F_{\chi_2^2(0)/2}^{-1}(\alpha), \dots, F_{\chi_K^2(0)/K}^{-1}(\alpha) \right)$ . Therefore, for  $\mu^2 \leq x$

$$P[Q > x^*] \leq \mathbb{P}[\chi_{n(x_1)}^2/n(x_1) > x_1(x, x^*)] \leq \alpha$$

Now assume that  $B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})$  is bounded in probability. Then  $c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}}))$  is bounded above in probability by some  $c^*$ . Then

$$\min \left[ \mathbb{P}\left(\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau)\right), \mathbb{P}\left(\widehat{F}_{eff} > c^*\right) \right] \xrightarrow{p} \mathbb{P}(F_{eff}^* > c^*) \quad (\text{A.29})$$

But then by the triangle inequality

$$\mathbb{P}(F_{eff}^* > c^*) \geq \mathbb{P}\left(\mu > \sqrt{c^*} + \sqrt{\sum_{i=1}^K c_i X_i^2}\right) \quad (\text{A.30})$$

where  $c_i$  are the eigenvalues of  $\mathbf{W}_2$  and  $X_i$  are iid standard normal. The right-hand side in (A.30) clearly converges to 1 as  $\mu^2 \rightarrow \infty$ , proving the second part of the Lemma.

## A.3 Proofs for Chapter 3

### A.3.1 Proof of Lemma 1

This lemma is used to show that the class of  $\alpha$ -conditionally similar-on-the-boundary tests and the class of  $\alpha$ -similar-on-the-boundary-tests is compact, relative to the space of essentially bounded measurable functions endowed with the weak\* topology (Lemma 1, below).

**PRELIMINARIES 1** ( $L^1$  and  $L^\infty$ ): Since the sample space  $\mathbf{X} \in \mathcal{B}(\mathbb{R}^s)$ , the triplet  $(\mathbf{X}, \mathcal{B}(\mathbb{R}^s)_{\mathbf{X}}, \lambda^s)$  is a well-defined  $\sigma$ -finite measure space. Note that  $\mathcal{B}(\mathbb{R}^s)_{\mathbf{X}} = \mathcal{B}(\mathbf{X})$  whenever  $\mathbf{X}$  is endowed with the sub-space topology relative to  $\mathbb{R}^s$ . Following Rudin (2006), p. 65, let  $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$  denote the space of all real-valued measurable functions  $f$  that satisfy  $\|f\|_1 \equiv \int_{\mathbf{X}} |f(x)| dx < \infty$ . Let  $L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$  denote the class of all essentially bounded real-valued measurable functions (Rudin (2006) p. 66).

**Remark 17.** Identify the class of all tests  $\mathcal{C}$  as a subset of  $L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$

$$\mathcal{C} \equiv \{\phi \in L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s) \mid \phi(x) \in [0, 1] \text{ for } \lambda^s\text{-a.e. } x \in \mathbf{X}\}.$$

And note that the elements of any statistical model  $\{f(x, \theta)\}_{\theta \in \Theta}$  are elements of  $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ , by the definition of probability density function  $\int_{\mathbf{X}} f(x, \theta) dx = 1 < \infty$  for all  $\theta \in \Theta$ .

**PRELIMINARIES 2** (The dual space of  $L^1$ ): Let  $[L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$  denote the dual space of  $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ , i.e., the space of all continuous (w.r.t.  $\|f\|_1$ ) linear functionals on  $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ ; see Rudin (2005), p. 56. Let  $\Lambda$  denote an element of the dual space  $[L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$ . By Theorem 6.16 in Rudin (2006), p. 127 and Theorem 1.18 in Rudin (2005), p. 15; the space  $[L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$  is isometrically isomorphic to

$L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ . Therefore, one can identify each functional  $\Lambda$  with a unique element (up to equivalence)  $g \in L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ , and vice versa: for  $f \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)^*$ ,  $\Lambda \in [L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$  is of the form

$$\Lambda(f) \equiv \int_{\mathbf{X}} g(x)f(x)dx \quad \text{for some } g \in L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s).$$

**PRELIMINARIES 3** (weak\* topology on  $L^\infty$ ): Endow the space  $L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$  with the topology induced by the weak\*-topology on the space  $[L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)]^*$ ; see Rudin (2005), p. 67, 68. The new topological space is denoted by  $(L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s), \mathcal{T}^*)$ . Denote convergence in such topology by  $\rightarrow^*$ . Note that, by definition,  $\{g_n\}_{n \in \mathbb{N}} \rightarrow^* g$  if and only if

$$\int_{\mathbf{X}} f(x)g_n(x)dx \rightarrow \int_{\mathbf{X}} f(x)g(x)dx \quad \forall f \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s).$$

Let  $(\mathbf{X}, \Theta, f, \Theta_0)$  be a hypothesis testing problem. Let  $\mathcal{G} \subset L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$  be an arbitrary collection of bounded functions. Define

$$\mathcal{C}(\alpha\text{-}\mathcal{G}) \equiv \{\phi \in \mathcal{C} \mid \mathbb{E}_\theta[(\phi(X) - \alpha)g(X)] = 0 \quad \forall \theta \in \text{Bd}\Theta \quad \forall g \in \mathcal{G}\}$$

Let  $(L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s), \mathcal{T}^*)$  be the space of essentially bounded functions topologized with the weak\* topology. For any  $A \subset L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ , let  $\mathcal{T}_A^*$  denote the subset topology induced by  $\mathcal{T}^*$

**LEMMA 1:** The set  $\mathcal{C}(\alpha\text{-}\mathcal{G})$  is compact relative to  $(\mathcal{C}, \mathcal{T}_\mathcal{C}^*)$ .

*Proof.* The outline of the proof is the following. I show that the set  $\mathcal{C}(\alpha\text{-}\mathcal{G})$  is a sequentially closed subset of  $\mathcal{C}$  with the relative weak\* topology. Then I use the Banach-Alaoglu theorem and the topological separability of  $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$  to establish the compactness of  $\mathcal{C}(\alpha\text{-}\mathcal{G})$ .

(*Sequential Closedness*) Take any convergent sequence of tests  $\phi_n \rightarrow^* \phi$  with  $\{\phi_n\}_{n \in \mathbb{N}} \subseteq \mathcal{C}(\alpha\text{-}\mathcal{G})$ . I want to show that  $\phi \in \mathcal{C}(\alpha\text{-}\mathcal{G})$ . First, I show that  $\phi(x) \in \mathcal{C}$ ; i.e.,  $\phi \in [0, 1]$  for almost every  $x \in \mathbf{X}$ . Suppose not. Then there exists a measurable set  $A \in \mathcal{B}(\mathbf{X})$  with  $\lambda^s(A) > 0$  such that  $\phi(x) > 1$  or  $\phi(x) < 0$  for all  $x \in A$ . Without loss of generality assume  $\phi(x) > 1$ . Since  $\lambda^s$  is  $\sigma$ -finite, there exists a countable collection  $\{E_n\}_{n \in \mathbb{N}}$  such that  $\cup_{n \in \mathbb{N}} E_n = \mathbf{X}$  and  $\lambda^s(E_n) < \infty$  for every  $n$ . Consider the sequence of sets  $\{A \cap E_n\}_{n \in \mathbb{N}}$ . Note that  $0 \leq \lambda^s(A \cap E_n) < \infty$  for all  $n \in \mathbb{N}$ . In addition, there exists  $N \in \mathbb{N}$  for which  $0 < \lambda^s(A \cap E_N)$ , otherwise  $\lambda^s(A) = \lambda^s(\cup_{n=1}^{\infty} (A \cap E_n)) \leq \sum_{n=1}^{\infty} \lambda^s(A \cap E_n) = 0$ . Consider the indicator function  $\mathbb{1}_{A \cap E_N}$ . Since  $0 < \lambda^s(A \cap E_N) < \infty$ , the indicator function  $\mathbb{1}_{A \cap E_N} \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ . Note that

$$\lambda^s(A \cap E_N) < \int_{\mathbf{X}} \mathbb{1}_{A \cap E_N}(x) \phi(x) dx = \lim_{n \rightarrow \infty} \int_{\mathbf{X}} \mathbb{1}_{A \cap E_N}(x) \phi_n(x) dx \leq \lambda^s(A \cap E_N).$$

A contradiction. Therefore  $\phi(x) \leq 1$   $\lambda^s$ -almost everywhere in  $\mathbf{X}$ . An analogous argument yields  $\phi(x) \geq 0$   $\lambda^s$ -almost everywhere. Therefore  $\phi \in \mathcal{C}$ . Now, I need to show that  $\phi \in \mathcal{C}(\alpha\text{-}\mathcal{G})$ . By assumption, for every  $\theta \in \text{Bd}\Theta_0$   $f(\cdot; \theta)$  is an element of  $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ . In addition,  $g \in \mathcal{G}$  is bounded. Consequently,  $f(\cdot, \theta)g(\cdot) \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ . Since  $\phi_n \in \mathcal{C}(\alpha\text{-}\mathcal{G})$  for every  $n \in \mathbb{N}$  weak\* convergence yields

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \int_{\mathbf{X}} f(x; \theta) g(x) (\phi_n(x) - \alpha) dx &= \left( \lim_{n \rightarrow \infty} \int_{\mathbf{X}} f(x; \theta) g(x) \phi_n(x) dx \right) \\ & &- \int_{\mathbf{X}} f(x; \theta) g(x) \alpha dx \\ &= \int_{\mathbf{X}} f(x; \theta) g(x) \phi(x) dx - \int_{\mathbf{X}} f(x; \theta) g(x) \alpha dx \\ &= \int_{\mathbf{X}} f(x; \theta) g(x) (\phi(x) - \alpha) dx. \end{aligned}$$

So  $\phi \in \mathcal{C}(\alpha\text{-}\mathcal{G})$ . This implies  $\mathcal{C}(\alpha\text{-}\mathcal{G})$  is sequentially closed in  $\mathcal{C}$  endowed with the weak\* topology.



(*Compactness*) Let

$$V \equiv \left\{ f \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s) : \int_{\mathbf{X}} |f(x)| dx \leq 1 \right\}$$

Note that  $V$  is a neighborhood of the function  $\mathbf{0}$  in the space  $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ . Let

$$K \equiv \left\{ g \in L^\infty(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s) : \left| \int_{\mathbf{X}} f(x)g(x) dx \right| \leq 1 \quad \forall \quad f \in V \right\}. \quad (\text{A.31})$$

Note that  $C(\alpha\text{-}\mathcal{G}) \subseteq \mathcal{C} \subseteq K$ , as for any test  $\left| \int_{\mathbf{X}} f(x)\phi(x) dx \right| \leq \int_{\mathbf{X}} |f(x)|\phi(x) dx \leq \int_{\mathbf{X}} |f(x)| dx \leq 1$ . By the Banach-Alaouglu Theorem the set  $K$  is compact in the weak\* topology; see Rudin (2005), p. 68, Theorem 3.15. Furthermore, the space  $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$  is topologically separable as  $(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$  is a separable measure space; see exercise 10, Chapter 1 of Stein (2011). Therefore, Theorem 3.16 in Rudin (2005) p. 70 implies that the topological space  $(K, \mathcal{T}_K^*)$  is compact and metrizable. Since every metrizable space is first-countable—consequently, Frechet-Urysohn—the sequential closure of  $C(\alpha)$  coincides with its closure. Therefore, the set  $D^*(\alpha)$  is a closed subset of the compact topological space  $(K, \mathcal{T}_K^*)$ . I conclude that  $(\mathcal{C}(\alpha\text{-}\mathcal{G}), \mathcal{T}_{\mathcal{C}(\alpha\text{-}\mathcal{G})}^*)$  is compact and metrizable.  $\square$

**COROLLARY 1:** The space of  $\alpha$ -similar-on-the-boundary tests,  $\mathcal{C}(\alpha\text{-sb})$  is weak\* compact.

*Proof.* Set  $\mathcal{G} = \{g(x) = 1 \quad \forall x \in \mathbf{X}\}$ .  $\square$

**COROLLARY 2:** The space of  $\alpha$ -conditionally similar-on-the boundary tests,  $\mathcal{C}(\alpha\text{-csb})$  is weak\* compact.

*Proof.* Set  $\mathcal{G} = \{g \mid g(x_1, x_2) = 1 \text{ if } x_2 \in \mathcal{F}; g(x_1, x_2) = 0 \text{ if } x \in \mathcal{F} \text{ for some } \mathcal{F} \in \mathcal{B}(\mathbf{X}_2)\}$ .

Fix  $\theta \in \text{Bd}\Theta_0$  and consider the random variables  $\phi(X_1, X_2)$  and  $Y \equiv X_2$  defined on

the probability space  $(\mathbf{X}, \mathcal{B}(\mathbf{X}), P_\theta)$ , where  $P_\theta$  is the measure induced by  $f(x; \theta)$ . Note that

$$\mathbb{E}_\theta[(\phi(X_1, X_2) - \alpha)g(X_1, X_2)] = \alpha \quad \forall g \in \mathcal{G}$$

implies that

$$\int_{\mathbf{X}_1 \times \mathcal{F}} (\phi(X_1, X_2) - \alpha) dP_\theta = 0 \quad \forall \mathcal{F} \in \mathcal{B}(\mathbf{X}_2).$$

By definition of conditional expectation (Billingsley (1995) p. 445), it follows that

$$E[\phi(X_1, X_2) - \alpha | X_2] = 0,$$

except, perhaps, in a set of measure zero under  $P_\theta$ . And this holds for every  $\theta \in \text{Bd}\Theta_0$ .  $\square$

### A.3.2 Proof of Theorem 1

**T1a (Outline):** I have shown that the class of tests  $\mathcal{C}(\alpha, \mathcal{G})$  is weak\* compact. This class is non-empty, as it contains the randomized test  $\phi(x) = \alpha$ . To establish Theorem 1 it will be sufficient to show that the objective function

$$\mathcal{W}^*(\phi) \equiv \mathbf{W} \left( \int_{\text{Int } \Theta_1} R(\phi, \theta) p_1(\theta) d\theta, \int_{\text{Int } \Theta_0} R(\phi, \theta) p_0(\theta) d\theta \right).$$

is continuous in the weak\* topology.

**T1a-Step 1** (Fubini's Theorem:) Since the image of any test  $\phi \in \mathcal{C}$  is contained in the interval  $[0, 1]$   $\lambda^s$ -a.e. and  $f(x; \theta) \in L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$  for all  $\theta$ , it follows that  $\left( \int_{\mathbf{X}} \phi(x) f(x; \theta) dx \right) \leq 1$  for every  $\theta \in \Theta$ . Furthermore, since  $p_1(x)$  and  $p_0(x)$  are also probability density functions on  $\text{Int}\Theta_1$  and  $\text{Int}\Theta_0$  the following holds

$$\int_{\text{Int}\Theta_1} \left( \int_{\mathbf{X}} \phi(x) f(x; \theta) dx \right) p_1(\theta) d\theta \leq 1 < \infty$$

and

$$\int_{\text{Int}\Theta_0} \left( \int_{\mathbf{X}} \phi(x) f(x; \theta) dx \right) p_0(\theta) d\theta \leq 1 < \infty.$$

Therefore, an application of Fubini's theorem in Billingsley (1995), p. 234 yields

$$\int_{\text{int}\Theta_1} R(\phi, \theta) p_1(\theta) d\theta \equiv \int_{\text{int}\Theta_1} \left( \int_{\mathbf{X}} (1 - \phi(x)) f(x; \theta) dx \right) p_1(\theta) d\theta = \int_{\mathbf{X}} (1 - \phi(x)) f_1^*(x) dx$$

and

$$\int_{\text{Int}\Theta_0} R(\phi, \theta) p_0(\theta) d\theta \equiv \int_{\text{int}\Theta_0} \left( \int_{\mathbf{X}} \phi(x) f(x; \theta) dx \right) p_0(\theta) d\theta = \int_{\mathbf{X}} \phi(x) f_0^*(x) dx.$$

where  $f_1^*$  and  $f_0^*$  are the "integrated" likelihoods given by

$$f_1^*(x) \equiv \int_{\text{int}\Theta_1} f(x; \theta) p_1(\theta) d\theta, \quad f_0^*(x) \equiv \int_{\text{int}\Theta_0} f(x; \theta) p_0(\theta) d\theta. \quad (\text{A.32})$$

Note that both  $f_1^*$  and  $f_0^*$  are elements of  $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda^s)$ . Note that the mapping

$\mathbf{W} : \mathbb{R}^2 \rightarrow \mathbb{R}$  induces a functional  $\mathcal{W}^*$  over  $\mathcal{D}$ :

$$\mathcal{W}^*(\phi) \equiv \mathbf{W} \left( \int_{\mathbf{X}} (1 - \phi(x)) f_1^*(x) dx, \int_{\mathbf{X}} \phi(x) f_0^*(x) dx \right) \quad (\text{A.33})$$

**T1a-Step 2** (Sequential Continuity of  $\mathcal{W}^*$ .) I now show that  $\mathcal{W}^*$  is continuous on the compact metrizable space  $(\mathcal{C}(\alpha\mathcal{G}), \mathcal{T}_{\mathcal{C}(\alpha\mathcal{G})}^*)$ . It suffices to establish sequential continuity. Take any sequence of tests  $\phi_n \rightarrow^* \phi$ . Since both  $f_1^*$  and  $f_0^*$  are elements of  $L^1(\mathbf{X}, \mathcal{B}(\mathbf{X}), \lambda_{\mathbf{X}})$ , convergence in the weak\* topology yields

$$\int_{\mathbf{X}} \phi_n(x) f_1^*(x) dx \rightarrow \int_{\mathbf{X}} \phi(x) f_1^*(x) dx \quad \text{and} \quad \int_{\mathbf{X}} \phi_n(x) f_0^*(x) dx \rightarrow \int_{\mathbf{X}} \phi(x) f_0^*(x) dx.$$

Therefore, the continuity of  $\mathbf{W}$  implies

$$\begin{aligned} \mathcal{W}^*(\phi_n) &\equiv \mathbf{W} \left( 1 - \int_{\mathbf{X}} \phi_n(x) f_1^*(x) dx, \int_{\mathbf{X}} \phi_n(x) f_0^*(x) dx \right) \\ &\rightarrow \mathbf{W} \left( 1 - \int_{\mathbf{X}} \phi(x) f_1^*(x) dx, \int_{\mathbf{X}} \phi(x) f_0^*(x) dx \right) \\ &= \mathcal{W}^*(\phi). \end{aligned}$$

Therefore,  $\mathcal{W}^*$  is a continuous functional defined on the compact space  $(\mathcal{C}(\alpha\text{-}\mathcal{G}), \mathcal{T}_{\mathcal{C}(\alpha\text{-}\mathcal{G})}^*)$ , and  $\mathcal{C}(\alpha\text{-}\mathcal{G}) \neq \emptyset$ , as it contains the test  $\phi(x) = \alpha$ . This implies  $M(\mathbf{W}, p_1, p_0, \mathcal{G}) \neq \emptyset$ .

**T1b :** Let  $\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G})$ . I show that if  $\phi' \in \mathcal{C}(\alpha\text{-}\mathcal{G})$  satisfies

$$\mathbb{E}_\theta[\phi'(X)] \leq \mathbb{E}_\theta[\phi^*(X)] \quad \forall \quad \theta \in \Theta_0 \quad (\text{A.34})$$

and

$$\mathbb{E}_\theta[\phi'(X)] \geq \mathbb{E}_\theta[\phi^*(X)] \quad \forall \quad \theta \in \Theta_1 \quad (\text{A.35})$$

then

$$\mathbb{E}_\theta[\phi'(x)] = \mathbb{E}_\theta[\phi^*(x)] \quad \forall \quad \theta \in \Theta = \Theta_0 \cup \Theta_1. \quad (\text{A.36})$$

Consequently, there is no test  $\phi' \in \mathcal{C}(\alpha\text{-}\mathcal{G})$  that “weakly dominates”  $\phi^*$  ; i.e,  $R(\phi', \theta) \leq R(\phi^*, \theta)$  with strict inequality for some  $\theta$ .

Suppose (A.34) and (A.35) hold, but (A.36) does not. Then, one of the following claims is true:

C1 There exists  $\tilde{\theta} \in \Theta_1$  such that  $\Delta_{\phi', \phi^*}(\tilde{\theta}) \equiv \mathbb{E}_{\tilde{\theta}}[\phi'(X)] - \mathbb{E}_{\tilde{\theta}}[\phi^*(X)] > 0$

C2 There exists  $\tilde{\theta} \in \Theta_0$  such that  $\Delta_{\phi', \phi^*}(\tilde{\theta}) \equiv \mathbb{E}_{\tilde{\theta}}[\phi'(X)] - \mathbb{E}_{\tilde{\theta}}[\phi^*(X)] < 0$ .

Assume first that C1 holds. The continuity of  $\Delta_{\phi', \phi^*}(\cdot)$  at  $\tilde{\theta}$  implies the existence of an open neighborhood  $\tau_{\tilde{\theta}}$  for which  $\Delta_{\phi', \phi^*}(\theta) < 0$  for all  $\theta \in \tau_{\tilde{\theta}}$ . Note that  $\Theta_1 \neq \emptyset$  is an open set. It follows that the set  $\mathcal{S}_{\tilde{\theta}}$  defined by  $\mathcal{S}_{\tilde{\theta}} \equiv \tau_{\tilde{\theta}} \cap \Theta_1$  satisfies three properties: it is non-empty, it is open, and it is contained in  $\Theta_1$ . Since  $p_1(\theta)$  has full support on  $\text{Int}\Theta_1$ ,  $\int_A p_1(\theta)d\theta > 0$  for any open set  $A$  contained in  $\Theta_1$ . Note that  $\Delta_{\phi', \phi^*}(\theta) > 0$  for

all  $\theta \in S_{\tilde{\theta}}$  and equations (A.34)-(A.35) imply

$$\int_{\text{Int}\Theta_0} \left( \int_{\mathbf{X}} \phi'(x) f(x; \theta) dx \right) p_0(\theta) d\theta \leq \int_{\text{Int}\Theta_0} \left( \int_{\mathbf{X}} \phi^*(x) f(x; \theta) dx \right) p_0(\theta) d\theta.$$

and

$$\int_{\Theta_1} \left( \int_{\mathbf{X}} (1 - \phi'(x)) f(x; \theta) dx \right) p_1(\theta) d\theta < \int_{\Theta_1} \left( \int_{\mathbf{X}} (1 - \phi^*(x)) f(x; \theta) dx \right) p_1(\theta) d\theta$$

The monotonicity of  $\mathbf{W}$  implies that  $\mathcal{W}^*(\phi') < \mathcal{W}(\phi^*)$ . This contradicts the fact that  $\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G})$ . I conclude C1 cannot hold.

Now, suppose C2 holds. Since the function  $g^*(x) = 1$  belongs to  $\mathcal{G}$ , then  $\tilde{\theta}$  must belong to  $\text{Int}\Theta_0$ . If  $\text{Int}\Theta_0 = \emptyset$  the proof is over. If  $\text{Int}\Theta_0 \neq \emptyset$  then —by analogy with the previous paragraph— there exists an open set  $S_{\tilde{\theta}}$  contained in  $\text{Int}\Theta_0$  such that  $\Delta_{\phi', \phi^*}(\theta) < 0$  for all  $\theta \in S_{\tilde{\theta}}$ . Since this set has positive probability under  $p_0$ , this implies

$$\int_{\text{Int}\Theta_0} \left( \int_{\mathbf{X}} \phi'(x) f(x; \theta) dx \right) p_0(\theta) d\theta < \int_{\text{Int}\Theta_0} \left( \int_{\mathbf{X}} \phi^*(x) f(x; \theta) dx \right) p_0(\theta) d\theta$$

and

$$\int_{\Theta_1} \left( \int_{\mathbf{X}} (1 - \phi'(x)) f(x; \theta) dx \right) p_1(\theta) d\theta \leq \int_{\Theta_1} \left( \int_{\mathbf{X}} (1 - \phi^*(x)) f(x; \theta) dx \right) p_1(\theta) d\theta.$$

Which, once again, contradicts the fact that  $\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G})$ .

Therefore, (A.34) and (A.35) imply (A.36). I conclude that  $\phi^*$  is admissible in  $\mathcal{C}(\alpha\text{-}\mathcal{G})$ .

**T1c (Outline):** Let  $\mathcal{G}^* \equiv \{g : \mathbf{X} \rightarrow \mathbb{R} \mid g(x) = 1 \quad \forall x \in \mathbf{X}\}$ , so that the class  $\mathcal{C}(\alpha\text{-}\mathcal{G}^*)$  coincides with  $\mathcal{C}(\alpha\text{-sb})$ . I show that a test  $\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G}^*)$  is admissible in the class of all tests. The proof is divided into two steps.

**STEP 1:** First I show that if  $\phi' \in \mathcal{C}$  satisfies

$$\mathbb{E}_\theta[\phi'(X)] \leq \mathbb{E}_\theta[\phi^*(X)] \quad \forall \quad \theta \in \Theta_0 \quad (\text{A.37})$$

and

$$\mathbb{E}_\theta[\phi'(X)] \geq \mathbb{E}_\theta[\phi^*(X)] \quad \forall \quad \theta \in \Theta_1 \quad (\text{A.38})$$

then  $\phi'$  is  $\alpha$ -similar on  $\text{Bd}\Theta_0$ . Consequently, any test  $\phi'$  that “weakly dominates”  $\phi^*$  (i.e,  $R(\phi', \theta) \leq R(\phi^*, \theta)$  with strict inequality for some  $\theta$ ) must be  $\alpha$ -similar on the boundary of  $\Theta_0$ .

Let  $\mathcal{C}_{ns} \subset \mathcal{C}$  be the class of tests that are not similar on the boundary of  $\Theta_0$ . This is,  $\phi \in \mathcal{C}_{ns}$  if and only if there exists  $\theta, \theta' \in \text{Bd}\Theta_0$  such that  $\mathbb{E}_\theta[\phi(x)] \neq \mathbb{E}_{\theta'}[\phi(x)]$ . Partition  $\mathcal{C}$  according to  $\mathcal{C}_{ns}$  so that  $\mathcal{C} \equiv \mathcal{C}_{ns} \cup (\mathcal{C} \setminus \mathcal{C}_{ns})$ . Take any test  $\phi' \in \mathcal{C}_{ns}$  that satisfies (A.34). Since  $\phi'$  is an element of  $\mathcal{C}_{ns}$  and  $\Theta_0$  contains its boundary (as it is closed), there exists  $\theta \in \text{Bd}\Theta_0$  such that  $\Delta_{\phi', \phi^*}(\theta) \equiv \mathbb{E}_\theta[\phi'(X)] - \mathbb{E}_\theta[\phi^*(X)] < 0$ . Because  $\Delta_{\phi', \phi^*}(\theta) < 0$  and the function  $\Delta_{\phi', \phi^*}(\cdot)$  is continuous at  $\theta$ , there exists an open neighborhood  $\tau_\theta \in \mathcal{T}$  such that  $\Delta_{\phi', \phi^*}(\theta) < 0$  for all  $\theta \in \tau_\theta$ . Since  $\theta$  is an element of  $\text{Bd}\Theta_0$ , then  $\tau_\theta \cap \Theta_1 \neq \emptyset$ . The latter implies there exists  $\theta_1 \in \Theta_1$  such that  $\Delta_{\phi', \phi^*}(\theta_1) = \mathbb{E}_{\theta_1}[\phi'(X)] - \mathbb{E}_{\theta_1}[\phi^*(X)] < 0$ . Therefore, equation (A.34) and (A.35) cannot hold. We conclude there is no test  $\phi' \in \mathcal{C}_{ns}$  that satisfies (A.34) and (A.35). Since  $\mathcal{C}_{ns}$  partitions  $\mathcal{C}$ , a test  $\phi' \in \mathcal{C}$  that satisfies (A.34) and (A.35) must be an element

of  $\mathcal{C} \setminus \mathcal{C}_{ns}$  (as  $\phi' \notin \mathcal{C}_{ns}$ ). Equation (A.34) implies  $\phi'$  is  $\alpha'$ -similar on the boundary with  $\alpha' \leq \alpha$ . Two cases follow:  $\alpha' < \alpha$  or  $\alpha' = \alpha$ . In the first case, the argument in the previous paragraph implies that  $\phi'$  will violate (A.35). We conclude that any test that satisfies (A.34) and (A.35) must be  $\alpha$ -similar on  $\text{Bd}\Theta_0$ .

**STEP 2:** Since  $\phi^* \in M(\mathbf{W}, p_1, p_0, \mathcal{G}^*)$ ,  $\phi^*$  is admissible in  $\mathcal{C}(\alpha\text{-}\mathcal{G}^*)$ . Therefore, there is no  $\alpha$ -similar-on-the-boundary test such that  $R(\phi', \theta) \leq R(\phi^*, \theta)$  with strict inequality for some  $\theta \in \Theta$ . Since —by Step 1— any test  $\phi' \in \mathcal{C}$  that satisfies (A.34) and (A.35) must be  $\alpha$ -similar on  $\text{Bd}\Theta_0$ , I conclude  $\phi^*$  is admissible in  $\mathcal{C}$

### A.3.3 Proof of Theorem 2

**STEP 1 T2:** (ECS-tests objective function). Let

$$\mathbf{X}_1(x_2) \equiv \{x_1 \in \mathbf{X}_1 \mid (x_1, x_2) \in \mathbf{X}\}.$$

Fubini's theorem (T1a-Step 1) implies that  $\phi^*$  is an ECS tests if and only if  $\phi^*$  solves the problem:

$$\min_{\phi \in \mathcal{C}} \quad \tau \int_X (1 - \phi(x)) f_1^*(x) dx + (1 - \tau) \int_X \phi(x) f_0^*(x) dx$$

subject to

$$\int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) f_{\text{Bd}}(x_1 | x_2) dx_1 = \alpha$$

except, perhaps, for  $x_2$  that belong to a set of measure zero under every  $h(x_2, \theta)$ ,  $\theta \in \text{Bd}\Theta_0$ . Re-write the objective function as

$$\max_{\phi \in \mathcal{C}} \quad \tau \int_X \phi(x) f_1^*(x) dx - (1 - \tau) \int_X \phi(x) f_0^*(x) dx.$$

The product structure of  $\mathbf{X}$  and the linearity of the integral allows a further expansion of the previous equation:

$$\max_{\phi \in \mathcal{C}} \int_{X_2} \left( \int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) [\tau f_1^*(x_1, x_2) - (1 - \tau) f_0^*(x_1, x_2)] dx_1 \right) dx_2$$

**STEP 2 T2:** (Necessity) First I show that the test  $\phi^*(x_1, x_2)$  that rejects the null hypothesis whenever

$$[\tau f_1^*(x_1, x_2) - (1 - \tau) f_0^*(x_1, x_2)] / f_{\text{Bd}}(x_1 | x_2) > c(x_2; \alpha)$$

is an element of the set  $M(\tau, p_1, p_0)$ —provided  $c(x_2, \alpha)$  is defined as the  $(1-\alpha)$  quantile of the random variable  $z_{\text{ecs}}(X_1, x_2; p_1, p_0, \tau)$  for every  $x_2 \in \mathbf{X}_2$ . That is to say

$$c(x_2; \alpha) \equiv \arg \min_{q \in \mathbf{X}_1(x_2)} \mathbb{E}_{f_{\text{Bd}}(x_1 | x_2)} \left[ \rho_{(1-\alpha)} \left( z_{\text{ecs}}(x_1, x_2; p_1, p_0, \tau) - q \right) \right].$$

Note first that the Generalized Neyman Pearson Lemma in Ferguson (1967) p. 204 implies that for a fixed  $x_2$  the test  $\phi^*(x_1, x_2)$  solves the problem

$$\max_{\phi \in \mathcal{C}} \int_{\{x_1 \in \mathbf{X}_1 \mid (x_1, x_2) \in \mathbf{X}\}} \phi(x_1, x_2) [\tau f_1^*(x_1, x_2) - (1 - \tau) f_0^*(x_1, x_2)] dx_1$$

subject to

$$\int_{\mathbf{X}_1(x_2)} \phi(x_1, x_2) f_{\text{Bd}}(x_1 | x_2) dx_1 = \alpha.$$

Hence, to show that  $\phi^*(x_1, x_2) \in M(\tau, p_1, p_0)$  it only remains to prove that  $\phi^*(x_1, x_2)$  is measurable. That is,  $\phi^*(x_1, x_2) \in \mathcal{C}_{X_2}(\alpha\text{-csb})$ . Assumption R1 imply that  $\phi^*(x_1, x_2)$  is continuous in  $x_1$ , for every  $x_2$ . Assumption R2 imply that the test is measurable in  $x_2$ , for every  $x_1$ . Therefore,  $\phi^*(x_1, x_2)$  is a Carathéodory function, as defined in



Aliprantis and Border (2006), p. 153. Since the sample space  $\mathbf{X}$  is separable (by assumption) and metrizable (for it is a subset of a euclidean space), Lemma 4.5.1 in Aliprantis and Border (2006) p. 153 implies  $\phi^* : \mathbf{X} \rightarrow [0, 1]$  is measurable.

**STEP 3 T2** (Sufficiency) Now I show that ECS tests are equal to  $\phi^*$  almost everywhere in  $\mathbf{X}$ . Let  $\phi' \in M(\tau, p_1, p_0)$ . I claim there is no set of lebesgue  $\lambda_{\mathbf{X}_2}$ -positive measure in  $A \in \mathcal{B}(\mathbf{X}_2)$  such that for each  $x_2 \in A$ ,  $\phi^*(x_1, x_2) \neq \phi'(x_1, x_2)$  in a set of lebesgue  $\lambda_{\mathbf{X}_1}$ -positive measure in  $\mathbf{X}_1$ . Suppose this is not the case. The maximizer in Step 2 is unique almost surely in  $X_1(x_2)$ . Hence, for all  $x_2 \in A$

$$\begin{aligned} \int_{\mathbf{X}_1(x_2)} \phi^*(x_1, x_2) [\tau f_1^*(x_1, x_2) - (1 - \tau) f_0^*(x_1, x_2)] dx_1 &> \\ \int_{\mathbf{X}_1(x_2)} \phi'(x_1, x_2) [\tau f_1^*(x_1, x_2) - (1 - \tau) f_0^*(x_1, x_2)] dx_1 \end{aligned}$$

Since  $A$  has positive measure, Integrating over  $x_2$  yields

$$\begin{aligned} \int_{\mathbf{X}} \phi^*(x) [\tau f_1^*(x) - (1 - \tau) f_0^*(x)] dx &> \\ \int_{\mathbf{X}} \phi'(x) [\tau f_1^*(x) - (1 - \tau) f_0^*(x)] dx \end{aligned}$$

which contradicts the fact that  $\phi'$  is an ECS-test.

### A.3.4 Example 1: Boundary Sufficiency for the IV model

Let  $\phi_n(\cdot, \mu, W)$  denote the probability density function of a multivariate normal random vector of dimension  $n$  with mean parameter  $\mu$  and covariance matrix  $W$ . Consider the standardized OLS coefficients:

$$\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \equiv (C_0 \otimes (Z'Z)^{1/2}) \hat{\gamma}_{\text{OLS}} = \begin{pmatrix} (Z'Z)^{-1/2} Z'(y_1 - Y_2 \beta_0) (b_0' \Omega b_0)^{-1/2} \\ \text{vec}[(Z'Z)^{1/2} Z'Y \Omega^{-1} A_0 (A_0' \Omega^{-1} A_0)^{-1/2}] \end{pmatrix}$$

which yield the statistical model for IV:

$$\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \sim \mathcal{N}_{k+nk} \left( \begin{pmatrix} (b_0' \Omega b_0)^{-1/2} (Z'Z)^{1/2} \Pi (\beta - \beta_0) \\ \text{vec}[(Z'Z)^{1/2} \Pi A' \Omega^{-1} A_0 (A_0' \Omega^{-1} A_0)^{-1/2}] \end{pmatrix}, \mathbb{I}_{k+nk} \right)$$

Note that

$$f(\gamma_1^*, \gamma_2^*; \beta = \beta_0, \Pi) = \phi_k(\gamma_1^*, \mathbf{0}, \mathbb{I}_k) \phi_{nk}(\gamma_2^*, \text{vec}[(Z'Z)^{1/2} \Pi A' \Omega^{-1} A_0 (A_0' \Omega^{-1} A_0)^{-1/2}], \mathbb{I}_{nk})$$

Therefore,  $\gamma_2^*$  is a boundary-sufficient statistic. Furthermore, for parameters  $(\beta, \Pi)$  on the boundary of the null hypothesis; that is, whenever  $\beta = \beta_0$  and  $\Pi \in \mathbb{R}^{nk \times n}$ :

$$\gamma_2^* \sim \mathcal{N}_{nk} \left( \text{vec}[(Z'Z)^{1/2} \Pi (A_0' \Omega^{-1} A_0)^{1/2}], \mathbb{I}_{nk} \right)$$

Using the properties of the kronecker product, the mean vector can be written as

$$((A_0' \Omega^{-1} A_0)^{1/2} \otimes (Z'Z)^{1/2}) \text{vec}(\Pi)$$

Since both matrices  $A_0' \Omega^{-1} A_0$  and  $(Z'Z)^{1/2}$  are of full rank the distributions for  $\gamma_2^*$  can be re-parameterized as a function of an unrestricted element of  $\mathbb{R}^{k \times n}$ . That  $\gamma_2^*$  is boundedly complete follows from Theorem 4.3.1 in Lehmann and Romano (2005) implies.

### A.3.5 Proof of Result 1

The rotated and standardized just-identified IV model has the following distribution:

$$\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \equiv (C_0 \otimes (Z'Z)^{1/2}) \hat{\gamma}_{\text{OLS}} \sim \mathcal{N}_{n+n^2} \left( \begin{pmatrix} C_0 \otimes (Z'Z)^{1/2} & \Pi\beta \\ & \text{vec}(\Pi) \end{pmatrix}, \mathbb{I}_{n+n^2} \right)$$

Consider the following prior distribution over the parameters  $(\beta, \Pi)$  :

$$\Pi = [\gamma_{21}, \gamma_{22}, \dots, \gamma_{2n}]$$

$$\beta = [\gamma_{21}, \gamma_{22}, \dots, \gamma_{2n}]^{-1} \gamma_1$$

where

$$\gamma \sim \mathcal{N}_{n+n^2}(\mathbf{0}, \lambda^2 \Omega \otimes (Z'Z)^{-1})$$

$\lambda^2$  is introduced as an extra parameter controlling the precision of the prior. I will show that regardless of the value of  $\lambda^2$  the ECS test rejects for large values of the Anderson and Rubin (1949) test.

Write  $\gamma = (\gamma'_1, \gamma'_{21}, \gamma'_{22}, \dots, \gamma'_{2n})'$  with  $\gamma_1$  is  $n \times 1$  and  $\gamma_{2i}$  is  $n \times 1$  for all  $i = 1, \dots, n$  and re-write the density of the vector  $(\gamma_1^*, \gamma_2^{*'})$  as  $f(\gamma_1^*, \gamma_2^{*'}; \beta(\gamma), \Pi(\gamma))$ . The integrated

likelihood in Theorem 2 is given by:

$$\begin{aligned}
f_1^*(\gamma_1^*, \gamma_2^*) &= \int_{\mathbb{R}^{n+n^2}} f(\gamma_1^*, \gamma_2^*; \beta(\gamma), \Pi(\gamma)) \phi_{n+n^2}(\gamma, \mathbf{0}, \lambda^2 \Omega \otimes (Z'Z)^{-1}) d\gamma \\
&= a_1 \int_{\mathbb{R}^{n+n^2}} \exp \left( -\frac{1}{2} \left[ \gamma^* - (C_0 \otimes (Z'Z)^{1/2}) \gamma \right]' \left[ \gamma^* - (C_0 \otimes (Z'Z)^{1/2}) \gamma \right] \right) \\
&\quad \phi_{n+n^2}(\gamma, \mathbf{0}, \lambda^2 \Omega \otimes (Z'Z)^{-1}) d\gamma \\
&\quad \left( \text{(where I have used the fact } \begin{pmatrix} \Pi(\gamma) \beta(\gamma) \\ \text{vec}(\Pi(\gamma)) \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \text{vec}(\gamma_{21}, \gamma_{22}, \dots, \gamma_{2n}) \end{pmatrix} \right) \\
&= a_2 \int_{\mathbb{R}^{n+n^2}} \exp \left( -\frac{1}{2} \left[ \gamma^* - (C_0 \otimes (Z'Z)^{1/2}) \gamma \right]' \left[ \gamma^* - (C_0 \otimes (Z'Z)^{1/2}) \gamma \right] \right) \\
&\quad \exp \left( -\frac{1}{2\lambda^2} \gamma' (\Omega^{-1} \otimes Z'Z) \gamma \right) d\gamma \\
&\quad \text{(By definition of } \phi_{n+n^2}) \\
&= a_2 \int_{\mathbb{R}^{n+n^2}} \exp \left( -\frac{1}{2} \left[ \gamma^* - (C_0 \otimes (Z'Z)^{1/2}) \gamma \right]' \left[ \gamma^* - (C_0 \otimes (Z'Z)^{1/2}) \gamma \right] \right) \\
&\quad \exp \left( -\frac{1}{2\lambda^2} \gamma' (C'_0 \otimes Z'Z^{1/2}) (C_0 \otimes (Z'Z)^{1/2}) \gamma \right) d\gamma \\
&\quad \text{(since } C'_0 C_0 = \Omega^{-1}) \\
&= a_3 \int_{\mathbb{R}^{n+n^2}} \exp \left( -\frac{1}{2} \left[ \gamma^* - \mu^* \right]' \left[ \gamma^* - \mu^* \right] \right) \exp \left( -\frac{1}{2\lambda^2} \mu^{*'} \mu^* \right) d\gamma \\
&= \left( \text{using the (linear) change of variables } \mu^* = (C_0 \otimes (Z'Z)^{1/2}) \gamma \right) \\
&\quad a_3 \left[ \int_{R^{m_1}} \exp \left( -\frac{1}{2} (\gamma_1^* - \mu_1^*)' (\gamma_1 - \mu_1^*) \right) \exp \left( -\frac{1}{2\lambda^2} \mu_1^{*'} \mu_1^* \right) d\mu_1^* \right] \\
&\quad \left[ \int_{R^{m_2}} \exp \left( -\frac{1}{2} (\gamma_2^* - \mu_2^*)' (\gamma_2 - \mu_2^*) \right) \exp \left( -\frac{1}{2\lambda^2} \mu_2^{*'} \mu_2^* \right) d\mu_2^* \right] \\
&= a_3 \exp \left( -\frac{1}{2} \gamma_1^{*'} \gamma_1^* \right) \int_{\mathbb{R}^{m_1}} \exp \left( \gamma_1^{*'} \mu_1^* \right) \exp \left( -\frac{1}{2b^2} \mu_1^{*'} \mu_1^* \right) d\mu_1^* \\
&\quad \exp \left( -\frac{1}{2} \gamma_2^{*'} \gamma_2^* \right) \int_{\mathbb{R}^{m_2}} \exp \left( x_2' \mu_2^* \right) \exp \left( -\frac{1}{2b^2} \mu_2^{*'} \mu_2^* \right) d\mu_2^* \\
&\quad \text{(where } b^2 = \lambda^2 / (1 + \lambda^2)) \\
&= c_1 \exp \left( -\frac{1}{2} \gamma_1^{*'} \gamma_1^* \right) \exp \left( \frac{b^2}{2} \gamma_1^{*'} \gamma_1^* \right) \exp \left( -\frac{1}{2} \gamma_2^{*'} \gamma_2^* \right) \exp \left( \frac{b^2}{2} \gamma_2^{*'} \gamma_2^* \right) \\
&\quad \text{(where I have used the definition of the Moment Generating Function} \\
&\quad \text{of a multivariate normal; and } c_1 \text{ is a non-negative constant.)}
\end{aligned}$$

Note that the boundary conditional likelihood for the IV model is given by:

$$f_{\text{Bd}}(\gamma_1^*|\gamma_2^*) = \phi_n(\gamma_1^*, \mathbf{0}, \mathbb{I}_n)$$

The ECS test statistic of Theorem 2 is defined as:

$$\begin{aligned} z(\gamma_1^*, \gamma_2^*; p_1) &= f_1^*(\gamma_1^*, \gamma_2^*)/f_{\text{Bd}}(\gamma_1^*|\gamma_2^*) \\ &= c_2 \exp\left(\frac{b^2}{2}\gamma_1^{*'}\gamma_1^*\right) \exp\left(-\frac{1}{2}\gamma_2^{*'}\gamma_2^*\right) \exp\left(\frac{b^2}{2}\gamma_2^{*'}\gamma_2^*\right) \end{aligned}$$

where  $c_2$  is a non-negative constant. Consider the critical value function

$$c(x_2; \alpha) = c_2 \exp\left(\frac{b^2}{2}\chi_{n,1-\alpha}^2\right) \exp\left(-\frac{1}{2}\gamma_2^{*'}\gamma_2^*\right) \exp\left(\frac{b^2}{2}\gamma_2^{*'}\gamma_2^*\right)$$

where  $\chi_{m_1,1-\alpha}^2$  is the  $1 - \alpha$  quantile of a central  $\chi_{m_1}^2$  distribution. Note that  $c(\gamma_2^*; \alpha)$  is measurable and for each fixed  $\gamma_2^*$  and

$$\mathbb{P}_{f_{\text{Bd}}(\gamma_1^*|\gamma_2^*)}(z(\gamma_1^*, \gamma_2^*; p_1) > c(\gamma_2^*, \alpha)) = \alpha$$

Hence, the assumptions of Theorem 2 are satisfied and the ECS rejects if:

$$c_2 \exp\left(\frac{b^2}{2}\gamma_1^{*'}\gamma_1^*\right) \exp\left(-\frac{1}{2}\gamma_2^{*'}\gamma_2^*\right) \exp\left(\frac{b^2}{2}\gamma_2^{*'}\gamma_2^*\right)$$

exceeds the critical value function

$$c_2 \exp\left(\frac{b^2}{2}\chi_{n,1-\alpha}^2\right) \exp\left(-\frac{1}{2}\gamma_2^{*'}\gamma_2^*\right) \exp\left(\frac{b^2}{2}\gamma_2^{*'}\gamma_2^*\right)$$

which happens if and only if  $\gamma_1^{*'}\gamma_1^* > \chi_{n,1-\alpha}^2$ . Since

$$\begin{pmatrix} \gamma_1^* \\ \gamma_2^* \end{pmatrix} \equiv (C_0 \otimes (Z'Z)^{1/2})\widehat{\gamma}_{\text{OLS}} = \begin{pmatrix} (Z'Z)^{-1/2}Z'(y_1 - Y_2\beta_0)(b_0'\Omega b_0)^{-1/2} \\ (Z'Z)^{1/2}Z'Y\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2} \end{pmatrix}$$

The ECS test rejects for large values of the Anderson and Rubin (1949) statistic.

### A.3.6 Proof of Result 2

Chamberlain's (2007) re-parameterization is given by:

$$\rho = (A'\Omega^{-1}A)^{1/2}(\Pi'Z'Z\Pi)^{1/2}, \quad \phi = C_0A/(A'\Omega^{-1}A)^{1/2}, \quad \omega = (Z'Z)^{1/2}\Pi/(\Pi'Z'Z\Pi)^{1/2}$$

where  $A \equiv [\beta, 1]'$ , and

$$C_0 \equiv \begin{pmatrix} (b_0'\Omega b_0)^{-1/2}b_0' \\ (A_0'\Omega^{-1}A_0)^{-1/2}A_0'\Omega^{-1} \end{pmatrix}$$

$$b_0 = [1, -\beta_0]', \quad A_0 = [\beta_0, 1]'$$

REMARK: The value  $\beta_0$  and the reduced-form covariance matrix  $\Omega$  impose a restriction on the possible values for the parameter  $\phi$ . Note first that  $\Omega C_0' C_0 = \mathbb{I}_2$ . Therefore:

$$\begin{aligned} (0, 1)\Omega C_0' \phi &= (0, 1)\Omega C_0' C_0 A / (A\Omega^{-1}A')^{1/2} \\ &= (0, 1)\mathbb{I}_2(\beta, 1)' / (A\Omega^{-1}A')^{1/2} \\ &= 1 / (A\Omega^{-1}A')^{1/2} \geq 0 \end{aligned}$$

Hence, the parameter  $\phi$  belongs to the intersection of the unit sphere  $\mathcal{S}^1$  and the half space  $\{x \in \mathbb{R}^2 \mid (0, 1)\Omega C_0' x \geq 0\}$ . In fact,

$$\begin{aligned}
(0,1)\Omega C'_0 &= \left( (0,1)\Omega b_0/(b'_0\Omega b_0)^{1/2}, (0,1)A_0/(A'_0\Omega^{-1}A_0)^{1/2} \right) \\
&= \left( (0,1)\Omega b_0/(b'_0\Omega b_0)^{1/2}, (\omega_1^2\omega_2^2 - \omega_{12}^2)^{1/2}/(b'_0\Omega b_0)^{1/2} \right) \\
&= \left( (0,1)\Omega b_0/(b'_0\Omega b_0)^{1/2}, (\omega_1^2\omega_2^2 - (\omega_{12} - \beta_0\omega^2 + \beta_0\omega_s^2)^2)^{1/2}/(b'_0\Omega b_0)^{1/2} \right) \\
&= \left( \omega_2 r(\beta_0), \omega_2 \sqrt{1 - r^2(\beta_0)} \right)
\end{aligned}$$

where  $r(\beta_0)$  corresponds to the structural correlation implied by  $\beta_0$ :

$$r(\beta_0) = (0,1)\Omega b_0 / \left[ (b'_0\Omega b_0)^{1/2} \omega_2 \right]$$

Thus, the domain for the parameter  $\phi$  in the canonical model is given by:

$$\Theta = \left\{ (\rho^2, \phi) \in \mathbb{R}_+ \times \mathcal{S}^1 : r(\beta_0)\phi_1 + \sqrt{1 - r^2(\beta_0)} \phi_2 \geq 0 \right\}$$

**DERIVATION OF THE INTEGRATED LIKELIHOODS:** Let:

$$f(S, T; \rho, \phi, \omega) = c_1 \exp \left( -\frac{1}{2} ([S', T']' - \rho(\phi \otimes \omega))' ([S', T']' - \rho(\phi \otimes \omega)) \right)$$

where  $c_1$  is a non-negative constant. Let  $Q \equiv [S, T]'[S, T]$ .

**STEP 1:** (Integrate  $\omega$ ) Note that:

$$\begin{aligned}
\tilde{f}(S, T; \rho, \phi) &\equiv c_2 \int_{\mathcal{S}^{k-1}} f(S, T; \rho, \phi, \omega) d\lambda_{\mathcal{S}^{k-1}}(\omega) \\
&= a_2(Q) \exp \left( -\rho^2/2 \right) \int_{\mathcal{S}^{k-1}} \exp \left( ([S, T]\phi)' \rho \omega \right) d\lambda_{\mathcal{S}^{k-1}}(\omega)
\end{aligned}$$

where  $\lambda_{\mathcal{S}^{k-1}}(\cdot)$  is the uniform measure over the  $k - 1$  dimensional sphere  $\mathcal{S}^{k-1}$  defined

in Chamberlain (2007) and Stroock (1999). In addition,

$$a_2(Q) \equiv c_2 \exp \left( -\frac{1}{2}[S'S + T'T] \right)$$

$c_2$  is a non-negative constant.

**STEP 2:** (Integrate  $\rho$ ) By assumption  $\rho \sim \sqrt{\lambda^2 \chi_k^2}$  independently of  $\phi$  and  $\omega$ . The latter implies that the density of  $\rho$  (with parameter  $\lambda^2$ ) is given by:

$$m_1(\rho; \lambda) \equiv \frac{1}{\lambda^2} \frac{1}{2^{k/2} \Gamma(k/2)} (\rho^2/\lambda^2)^{(k/2)-1} e^{-(\rho^2/2\lambda^2)} 2\rho$$

Note that using Fubini's Theorem and the change of variables formula:

$$\begin{aligned} & \int_{\mathbb{R}^+} \tilde{f}(S, T; \rho, \phi) m_1(\rho; \lambda^2) d\rho \\ &= a_2(Q) \int_{\mathbb{R}^+} \left( \exp(-\rho^2/2) \int_{S^{k-1}} \exp \left( \left[ (S, T) \rho \phi \right]' \omega \right) d\lambda_{S^{k-1}}(\omega) \right) m_1(\rho; \lambda) d\rho \\ &= a_2(Q) \int_{S^{k-1}} \left( \int_{\mathbb{R}^+} \exp \left( \left[ (S, T) \rho \phi \right]' \omega \right) m_1(\rho; \lambda) \exp(-\rho^2/2) d\rho \right) d\lambda_{S^{k-1}}(\omega) \\ &= a_3(Q) \int_{S^{k-1}} \left( \int_{\mathbb{R}^+} \exp \left( \left[ (S, T) \rho \phi \right]' \omega \right) \exp(-\rho^2/2b^2) \rho^{k-1} dp \right) d\lambda_{S^{k-1}}(\omega) \\ & \quad \text{(by definition of } m_1, b^2 \equiv [\lambda^2/(1 + \lambda^2)] \text{)} \\ &= a_3(Q) \int_{S^{k-1}} \left( \int_{\mathbb{R}^+} \exp \left( \left[ (S, T) \phi \right]' \rho \omega \right) \exp(-(\rho \omega)'(\rho \omega)/2b^2) \rho^{k-1} dp \right) d\lambda_{S^{k-1}}(\omega) \end{aligned}$$

where the last line follows from  $\omega' \omega = 1$  and  $a_3(Q) = a_2(Q) 2/(\lambda^k 2^{k/2} \Gamma(k/2))$ . Finally, consider the non-negative measurable function  $f : \mathbb{R}^k \rightarrow \mathbb{R}$

$$f(x) = \exp \left( \left[ (S, T) \phi \right]' x \right) \exp(-x'x/2b^2)$$

Theorem 5.2.2, p. 86 in Stroock (1999) implies:



$$\begin{aligned}
\int_{\mathbb{R}^+} \tilde{f}(S, T; \rho, \phi) m_1(\rho; \lambda^2) d\rho &= a_3(Q) \int_{\mathbb{R}^K} \exp\left(\left[(S, T)\phi\right]' x\right) \exp(-x'x/2b^2) dx \\
&= a_4(Q) \exp\left(\frac{b^2}{2} \phi' Q \phi\right)
\end{aligned}$$

where the last inequality follows by definition of the moment generating function of a  $k$ -dimensional multivariate normal evaluated at  $(S, T)\phi$ . Note that  $a_4(Q) \equiv (2\pi\lambda^2)^{k/2} a_3(Q)$ .

**STEP 3:** (Integrate  $\phi$ ) For simplicity, I will assume throughout the remaining part of this section that  $r(\beta_0) \geq 0$ . Consider the mapping  $m : [-\pi, \pi] \rightarrow \mathcal{S}^1$  given by  $m(\theta) = [-\sin(\theta), \cos(\theta)]$ . Note that  $m(\cdot)$  evaluated at  $-\pi$  gives the point  $(0, 1)$  in the unit circle. As  $\theta$  increases, the mapping  $m(\cdot)$  traces  $\mathcal{S}^1$  counter-clock wise. Therefore,

$$\mathcal{S}^1(r(\beta_0)) = \{\phi \in \mathcal{S}^1 \mid r(\beta_0)\phi_1 + \sqrt{1 - r^2(\beta_0)}\phi_2 \geq 0\}$$

can be expressed as:

$$\begin{aligned}
&\left\{ \theta \in [-\pi, \pi] : r(\beta_0)(-\sin(\theta)) + \sqrt{1 - r^2(\beta_0)} \cos(\theta) \geq 0 \right\} \\
&= \left[ \tan^{-1}\left(\sqrt{1 - r(\beta_0)^2}/r(\beta_0)\right) - \pi, \tan^{-1}\left(\sqrt{1 - r(\beta_0)^2}/r(\beta_0)\right) \right] \\
&= [\pi_0, \pi_0 + \pi], \quad \text{where} \quad \pi_0 \equiv \tan^{-1}\left(\sqrt{1 - r(\beta_0)^2}/r(\beta_0)\right) - \pi
\end{aligned}$$

and  $\pi_0 < 0$  when  $r(\beta_0) > 0$ . The parameter  $\phi \sim \mathcal{U}(\mathcal{S}^1(r(\beta_0)))$  if and only if  $\theta \sim \mathcal{U}[\pi_0, \pi_0 + \pi]$ . Define:

$$f_{[\pi_l, \pi_u]}^*(S, T) \equiv a_4(Q) \frac{1}{\pi_l - \pi_u} \int_{\pi_l}^{\pi_u} \exp\left(\frac{b^2}{2} \phi(\theta)' Q \phi(\theta) d\theta\right) d\theta; \quad \phi(\theta)' = [-\sin(\theta), \cos(\theta)]$$

The following Lemma is crucial for the derivation of the point and one-sided ECS tests in the IV model. Let

$$\zeta_{max} = \frac{1}{2} \left[ (S'S + T'T) + \sqrt{(S'S - T'T)^2 + 4(S'T)^2} \right]$$

$$\zeta_{min} = \frac{1}{2} \left[ (S'S + T'T) - \sqrt{(S'S - T'T)^2 + 4(S'T)^2} \right]$$

denote the maximum and minimum eigenvalues of the matrix  $Q \equiv [S, T]'[S, T]$  and let

**LEMMA 2:** . Let  $\pi_l, \pi_u$  belong to the interval  $[\pi_0, \pi + \pi_0]$ . Then

$$\begin{aligned} & f_{[\pi_l, \pi_u]}^*(S, T) \\ = & a_4(Q) \exp \left( \frac{b^2}{4} (\zeta_{max} + \zeta_{min}) \right) \frac{\pi}{\pi_u - \pi_l} I_0(\kappa(Q)) \left[ \Phi_{[0, 2\pi]}^{VM} \left( 2(\pi_u - \pi_0) | \kappa(Q), \mu(Q) \right) \right. \\ & \left. - \Phi_{[0, 2\pi]}^{VM} \left( 2(\pi_l - \pi_0) | \kappa(Q), \mu(Q) \right) \right] \end{aligned}$$

where  $\Phi_{[0, 2\pi]}^{VM}$  is the Von-Mises distribution in Mardia and Jupp (2000), p. 36 with *mean direction parameter*:

$$\mu(Q) = 2(\hat{\theta}_{max} - \pi_0) \geq 0$$

and *concentration parameter*

$$\kappa(Q) = \frac{b^2}{4} (\zeta_{max} - \zeta_{min}) \in [0, 2\pi]$$

$I_0(\cdot)$  is the modified Bessel function of the first kind, defined in Abramowitz and Stegun (1964), Section 9.6, p. 375.

*Proof.* Let  $L \equiv S'S - \zeta_{min}$ . Note that  $L$  is the Likelihood Ratio Statistic as defined in Andrews *et al.* (2006) p. 722. Define:

$$e_{max} \equiv \begin{cases} (L, S'T)' / \sqrt{L^2 + (S'T)^2} & \text{if } r(\beta_0)L + \sqrt{1 - r(\beta_0)}S'T > 0 \\ -(L, S'T)' / \sqrt{L^2 + (S'T)^2} & \text{if } r(\beta_0)L + \sqrt{1 - r(\beta_0)}S'T \leq 0 \end{cases}$$

Note that  $e_{max}$  is the maximum eigenvalue of the matrix  $Q$  adjusted to belong to the domain  $\mathcal{S}^1(r(\beta_0))$ . Define  $\hat{\theta} \in [\pi_0, \pi]$  implicitly by the following equation:

$$[-\sin(\hat{\theta}), \cos(\hat{\theta})]' = e_{max}$$

Therefore,

$$P \equiv \begin{pmatrix} -\sin(\hat{\theta}) & \cos(\hat{\theta}) \\ \cos(\hat{\theta}) & \sin(\hat{\theta}) \end{pmatrix}$$

yields the spectral decomposition of the matrix  $Q$ ; that is:

$$P \begin{pmatrix} \zeta_{max} & 0 \\ 0 & \zeta_{min} \end{pmatrix} P' = Q$$

. Note that for any  $\theta \in [\pi_l, \pi_u]$ :

$$P' \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} = \begin{pmatrix} \sin(\hat{\theta}) \sin(\theta) + \cos(\hat{\theta}) \cos(\theta) \\ -\cos(\hat{\theta}) \sin(\theta) + \sin(\hat{\theta}) \cos(\theta) \end{pmatrix} = \begin{pmatrix} \cos(\hat{\theta}_{max} - \theta) \\ \sin(\hat{\theta}_{max} - \theta) \end{pmatrix}$$

Therefore:

$$\begin{aligned}
f_{[\pi_l, \pi_u]}^*(S, T) &= a_4(Q) \frac{1}{\pi_u - \pi_l} \int_{\pi_l}^{\pi_u} \exp \left( \frac{b^2}{2} \left[ \zeta_{max} \cos^2(\hat{\theta}_{max} - \theta) + \zeta_{min} \sin^2(\hat{\theta}_{max} - \theta) \right] \right) d\theta \\
&= a_4(Q) \frac{1}{\pi_u - \pi_l} \int_{\hat{\theta}_{max} - \pi_u}^{\hat{\theta}_{max} - \pi_l} \exp \left( \frac{b^2}{2} \left[ \zeta_{max} \cos^2(\theta) + \zeta_{min} \sin^2(\theta) \right] \right) d\theta \\
&\quad \text{(where he have changed the integration variable)} \\
&= \exp \left( \frac{b^2}{2} \zeta_{min} \right) a_4(Q) \frac{1}{\pi_u - \pi_l} \int_{\hat{\theta}_{max} - \pi_u}^{\hat{\theta}_{max} - \pi_l} \exp \left( \frac{b^2}{2} \left[ (\zeta_{max} - \zeta_{min}) \cos^2(\theta) \right] \right) d\theta \\
&\quad \text{(as } \sin^2(\theta) + \cos^2(\theta) = 1) \\
&= a_4(Q) \exp \left( \frac{b^2}{2} \zeta_{min} \right) \exp \left( \frac{b^2}{4} (\zeta_{max} - \zeta_{min}) \right) \frac{1}{\pi_u - \pi_l} \\
&\quad \int_{\hat{\theta}_{max} - \pi_u}^{\hat{\theta}_{max} - \pi_l} \exp \left( \frac{b^2}{4} (\zeta_{max} - \zeta_{min}) \cos(2\theta) \right) d\theta \\
&\quad \text{(as } 2 \cos^2(\theta) - 1 = \cos(2\theta)) \\
&= a_4(Q) \exp \left( \frac{b^2}{2} \zeta_{min} \right) \exp \left( \frac{b^2}{4} (\zeta_{max} - \zeta_{min}) \right) \frac{1}{2(\pi_u - \pi_l)} \\
&= \int_{2(\hat{\theta}_{max} - \pi_u)}^{2(\hat{\theta}_{max} - \pi_l)} \exp \left( \kappa(Q) \cos(\theta) \right) d\theta \\
&\quad \text{(where we have used the change of variable } \tilde{\theta} = 2\theta) \\
&\quad \left( \kappa(Q) \equiv \frac{b^2}{4} (\zeta_{max} - \zeta_{min}) \right) \\
&= a_4(Q) \exp \left( \frac{b^2}{4} (\zeta_{max} + \zeta_{min}) \right) \frac{1}{2(\pi_u - \pi_l)} \\
&\quad \int_{2(\pi_l - \pi_0)}^{2(\pi_u - \pi_0)} \exp \left( \kappa(Q) \cos(\theta - 2(\hat{\theta}_{max} - \pi_0)) \right) du \\
&\quad \text{(where we have used } u = 2(\hat{\theta}_{max} - \pi_0) - \theta)
\end{aligned}$$

Note that  $\hat{\theta}_{max} - \pi_0 \in [0, \pi]$ . Therefore,  $\mu(Q) \equiv 2(\hat{\theta}_{max} - \pi_0) \in [0, 2\pi]$ . Using the definition of the Von-Mises distribution (supported on  $[0, 2\pi]$ ) in Mardia and Jupp

(2000) it follows that:

$$\begin{aligned}
& f_{[\pi_l, \pi_u]}^*(S, T) \\
= & a_4(Q) \exp\left(\frac{b^2}{4}(\zeta_{\max} + \zeta_{\min})\right) \frac{\pi}{\pi_u - \pi_l} I_0(\kappa(Q)) \left[ \Phi_{[0, 2\pi]}^{VM}\left(2(\pi_u - \pi_0)|\kappa(Q), \mu(Q)\right) \right. \\
& \left. - \Phi_{[0, 2\pi]}^{VM}\left(2(\pi_l - \pi_0)|\kappa(Q), \mu(Q)\right) \right]
\end{aligned}$$

$\mu(Q)$  is the mean direction parameter, and  $\kappa(Q)$  is the concentration parameter.  $I_0(\cdot)$  is the modified Bessel function of the first kind, defined in Abramowitz and Stegun (1964), Section 9.6, p. 375.  $\square$

**PROOF OF RESULT 2:** (ECS test for  $\phi_1 = 0$ ) From Lemma 2 above it follows that the integrated likelihood for independent priors:

$$\phi \sim \mathcal{U}(\mathcal{S}^1(r(\beta_0))) \quad \omega \sim \mathcal{U}(\mathcal{S}^k) \quad \rho \sim \sqrt{\lambda^2 \chi_k^2}$$

is given by:

$$f_1^*(S, T) = f_{[\pi_0, \pi_0 + \pi]}^* = a_1 \exp\left(-\frac{1}{2}[S'S + T'T]\right) \exp\left(\frac{b^2}{4}(S'S + T'T)\right) I_0(\kappa(Q))$$

where  $a_1$  is a non-negative constant,  $b^2 = \lambda^2/(1 + \lambda^2)$ , and

$$\kappa(Q) = \frac{b^2}{4} \left( (S'S - T'T)^2 + 4(S'T)^2 \right)^{1/2}$$

Note that the boundary conditional likelihood for the model is given by:

$$f_{\text{Bd}}(S, T) = a_2 \exp\left(-\frac{1}{2}S'S\right)$$

Both  $f_1^*(S, T)$  and  $f_{\text{Bd}}^*(S, T)$  are separately continuous, so that Assumption R1 is verified. Note that:

$$z(S, T, p_1) = \frac{a_1}{a_2} \exp \left( -\frac{1}{2} T' T \right) \exp \left( \frac{b^2}{4} [S' S + T' T] \right) I_0(\kappa(Q))$$

The quantile function  $c(T, \alpha)$  is continuous in  $T$  and, therefore, measurable. So that the ecs test rejects if

$$\frac{a_1}{a_2} \exp \left( -\frac{1}{2} T' T \right) \exp \left( \frac{b^2}{4} [S' S + T' T] \right) I_0(\kappa(Q)) \geq c(T, \alpha)$$

Which holds if and only if:

$$S' S - T' T + \frac{4(1 + \lambda^2)}{\lambda^2} \ln \left[ I_0 \left( \frac{\lambda^2}{4(1 + \lambda^2)} \left( (S' S - T' T)^2 + 4(S' T)^2 \right)^{1/2} \right) \right]$$

is larger than the critical value function  $c^*(T, \alpha)$ , defined as the  $1 - \alpha$  quantiles (conditional on  $T$ ) of the expression above under the distribution  $S \sim \mathcal{N}_k(\mathbf{0}, \mathbb{I}_k)$ .

### A.3.7 Proof of Result 2\*

Let:

$$\mathcal{S}^1(r(\beta_0))^- \equiv \{\phi \in \mathcal{S}^1(r(\beta_0)) \mid \phi_1 \leq 0\}$$

Define  $\mathcal{S}^1(r(\beta_0))^+$  analogously. Under the assumption  $r(\beta_0) \geq 0$ :

$$[-\sin(\theta), \cos(\theta)] \in \mathcal{S}^1(r(\beta_0))^- \iff \theta \in [0, \pi_0 + \pi]$$

Consider the independent priors under the alternative:

$$\phi \sim \mathcal{U}(\mathcal{S}^1(r(\beta_0))^+) \quad \omega \sim \mathcal{U}(\mathcal{S}^k) \quad \rho \sim \sqrt{\lambda^2 \chi_k^2}$$

and the independent priors under the null:

$$\phi \sim \mathcal{U}(\mathcal{S}^1(r(\beta_0))^-) \quad \omega \sim \mathcal{U}(\mathcal{S}^k) \quad \rho \sim \sqrt{\lambda^2 \chi_k^2}$$

RESULT 2\*: *The  $\alpha$ -ECS test for the problem  $\mathbf{H}_0 : \phi_1 \leq 0$  vs.  $\mathbf{H}_1 : \phi_1 > 0$  in an over-identified IV model with a single endogenous regressors and the priors above rejects if the statistic:*

$$\begin{aligned} & z(S, T; \tau) \\ \equiv & \exp\left(\frac{b^2}{4}(S'S - T'T)\right) I_0(\kappa(Q)) \\ & \left[ \frac{\tau}{\pi_0} \Phi_{[0, 2\pi]}^{VM}\left(-2\pi_0 \mid \kappa(Q), \mu(Q)\right) + \frac{1-\tau}{\pi_0 + \pi} \left(1 - \Phi_{[0, 2\pi]}^{VM}\left(-2\pi_0 \mid \kappa(Q), \mu(Q)\right)\right) \right] \end{aligned}$$

is smaller than the critical value function  $c_{os}^*(T; \lambda^2, \alpha)$ , defined as the  $(1 - \alpha)$  quantiles of  $z(S, T; \tau)$  with  $S \sim \mathcal{N}_k(\mathbf{0}, \mathbb{I}_k)$  and  $T$  fixed. The function  $I_0(\cdot)$  is the modified Bessel function of the first kind of order zero defined in Section 9.6, p. 375 of Abramowitz and Stegun (1964).

**PROOF OF RESULT 2\*:** From Lemma 2 it follows that the integrated likelihood under the alternative is given by:

$$\begin{aligned} f_1^*(S, T) = f_{[\pi_0, 0]}^* &= a_1 \exp\left(-\frac{1}{2}[S'S + T'T]\right) \exp\left(\frac{b^2}{4}(S'S + T'T)\right) \frac{\pi}{-\pi_0} I_0(\kappa(Q)) \\ & \quad \Phi_{[0, 2\pi]}^{VM}\left(-2\pi_0 \mid \kappa(Q), \mu(Q)\right) \end{aligned}$$

And the integrated likelihood under the null:

$$\begin{aligned} f_0^*(S, T) = f_{[0, \pi_0 + \pi]}^* &= a_1 \exp\left(-\frac{1}{2}[S'S + T'T]\right) \exp\left(\frac{b^2}{4}(S'S + T'T)\right) \frac{\pi}{\pi_0 + \pi} I_0(\kappa(Q)) \\ & \quad \left[1 - \Phi_{[0, 2\pi]}^{VM}\left(-2\pi_0 \mid \kappa(Q), \mu(Q)\right)\right] \end{aligned}$$

Therefore:

$$\begin{aligned}
& [\tau f_1^*(S, T) - (1 - \tau) f_0^*(S, T)] / f_{\text{Bd}}(S, T) \\
&= \frac{a_1}{a_2} \exp\left(-\frac{1}{2} T' T\right) \exp\left(\frac{b^2}{4} (S' S + T' T)\right) \pi I_0(\kappa(Q)) \\
&\quad \left[ -\frac{\tau}{\pi_0} \Phi_{[0, 2\pi]}^{VM}\left(-2\pi_0 \mid \kappa(Q), \mu(Q)\right) - \frac{1 - \tau}{\pi_0 + \pi} \left(1 - \Phi_{[0, 2\pi]}^{VM}\left(-2\pi_0 \mid \kappa(Q), \mu(Q)\right)\right) \right]
\end{aligned}$$

Therefore, the ECS one-sided test rejects if:

$$\begin{aligned}
z(S, T; \tau) \equiv & \exp\left(\frac{b^2}{4} (S' S - T' T)\right) I_0(\kappa(Q)) \\
& \left[ \frac{\tau}{\pi_0} \Phi_{[0, 2\pi]}^{VM}\left(-2\pi_0 \mid \kappa(Q), \mu(Q)\right) \right. \\
& \left. + \frac{1 - \tau}{\pi_0 + \pi} \left(1 - \Phi_{[0, 2\pi]}^{VM}\left(-2\pi_0 \mid \kappa(Q), \mu(Q)\right)\right) \right]
\end{aligned}$$

is smaller than the  $(1 - \alpha)$  quantile of the distribution of the statistic  $z(S, T; \tau)$  computed with  $T$  fixed and  $S \sim \mathcal{N}_k(\mathbf{0}, \mathbb{I}_k)$ .

### A.3.8 Proof of Lemma 3

Let  $W$  be a  $s \times s$ ,  $s \in \mathbb{N} \setminus \{1\}$ , symmetric matrix partitioned in the following blocks

$$W = \begin{pmatrix} W_1 & W_{12} \\ W_{21} & W_2 \end{pmatrix},$$

where  $W_1$  is  $s_1 \times s_1$  and  $W_2$  is  $s_2 \times s_2$ . Let

$$R_{12} = W_1^{-1/2} W_{12} W_2^{-1/2} \quad (s_1 \times s_2 \text{ matrix}),$$

where  $W_i^{-1/2}$  denotes the symmetric square root of  $W_i$ . Define  $B \equiv \mathbb{I}_{s_2} - R_{12}' R_{12}$ . Note



that  $B$  is the Schur complement of  $\mathbb{I}_{s_1}$  in the positive definite matrix:

$$\begin{pmatrix} \mathbb{I}_{s_2} & R'_{12} \\ R_{12} & \mathbb{I}_{s_1} \end{pmatrix}.$$

Consequently, there is a positive definite and symmetric matrix,  $B^{-1/2}$ , such that  $B^{-1/2}BB^{-1/2} = \mathbb{I}_{s_2}$ . Let

$$D = \begin{pmatrix} W_1^{-1/2} & \mathbf{0} \\ -B^{-1/2}R'_{12}W_1^{-1/2} & B^{-1/2}W_2^{-1/2} \end{pmatrix}.$$

**LEMMA 3:**  $DWD' = \mathbb{I}_s$

*Proof.* Note that

$$\begin{aligned} DW &= \begin{pmatrix} W_1^{-1/2} & \mathbf{0} \\ -B^{-1/2}R'_{12}W_1^{-1/2} & B^{-1/2}W_2^{-1/2} \end{pmatrix} \begin{pmatrix} W_1 & W_{12} \\ W_{21} & W_2 \end{pmatrix} \\ &= \begin{pmatrix} W_1^{1/2} & W_1^{-1/2}W_{12} \\ -B^{-1/2}R'_{12}W_1^{1/2} + B^{-1/2}W_2^{-1/2}W_{21} & -B^{-1/2}R'_{12}W_1^{-1/2}W_{12} + B^{-1/2}W_2^{1/2} \end{pmatrix} \end{aligned}$$

Therefore,  $DWD'$  equals

$$\begin{aligned} &\begin{pmatrix} W_1^{1/2} & W_1^{-1/2}W_{12} \\ -B^{-1/2}R'_{12}W_1^{1/2} + B^{-1/2}W_2^{-1/2}W_{21} & -B^{-1/2}R'_{12}W_1^{-1/2}W_{12} + B^{-1/2}W_2^{1/2} \end{pmatrix} \\ &\begin{pmatrix} W_1^{-1/2} & -W_1^{-1/2}R_{12}B^{-1/2} \\ \mathbf{0} & W_2^{-1/2}B^{-1/2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbb{I}_K & -R_{12}B^{-1/2} + R_{12}B^{-1/2} \\ -B^{-1/2}R'_{12} + B^{-1/2}R'_{12} & \mathbf{0} + -B^{-1/2}R'_{12}R_{12}B^{-1/2} + B^{-1/2}B^{-1/2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbb{I}_{s_1} & \mathbf{0} \\ \mathbf{0} & \mathbb{I}_{s_2} \end{pmatrix} \end{aligned}$$

□

### A.3.9 Proof of Result 3

The result is established in two steps. First, I use a simple change of variables formula that simplifies the derivation of the integrated likelihood. Second, I derive the ECS test.

**STEP 1:** Following the notation in Billingsley (1995) let

$$\Omega \equiv \mathbb{R}^{n+1}, \mathcal{F} \equiv \mathcal{B}(\mathbb{R}^{n+1}), \Omega' \equiv \mathbb{R}^{2m}, \mathcal{F}' \equiv \mathcal{B}(\mathbb{R}^{2m})$$

By assumption, the function  $T \equiv C^* : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{2m}$  is measurable. The prior  $p_1$  on  $\mathbb{R}^{n+1}$  and the function  $C^*$  induce a probability measure over the measurable space  $(\Omega', \mathcal{F}')$  in the usual way:

$$\mu T^{-1}(A') \equiv \mathbb{P}^*(A') \equiv \int_{\{x \in \mathbb{R}^{n+1} \mid C^*(x) \in A'\}} p_1(x) dx$$

Also, the measure  $P^*(A')$  is  $\mathcal{N}_{2m}(\mathbf{0}, \Omega(\theta_0))$ , by assumption. Define  $f : \mathbb{R}^{2m} \rightarrow \mathbb{R}$  by:

$$f(x) = c \exp \left( -\frac{1}{2} [\gamma(\theta_0) - D(\theta_0)x]' [\gamma(\theta_0) - D(\theta_0)x] \right).$$

Let  $\mu$  denote the probability measure associated with  $p_1$  in  $\mathbb{R}^{n+1}$ . Theorem 16.13 in Billingsley (1995) imply

$$\int_{\mathbb{R}^{n+1}} f(C^*(\theta^*, \delta)) \mu(d(\theta^*, \delta)) = \int_{\mathbb{R}^{2m}} f(x) \mu T^{-1}(dx),$$

which by Theorem 16.11 in Billingsley (1995) and the definition of a density (with respect to lebesgue measure) yield:

$$\int_{\mathbb{R}^{n+1}} f^*(C^*(\theta^*, \delta)) p_1(\theta^*, \delta) d\theta^* d\delta = \int_{\mathbb{R}^{2m}} f(x) \phi_{2m}(x, \mathbf{0}, \Omega(\theta_0)) dx.$$

Note that the integrated likelihood

$$\begin{aligned}
f_1^*(\gamma(\theta_0)) &= c \exp\left(-\frac{1}{2}\gamma(\theta_0)'\gamma(\theta_0)\right) \int_{\mathbb{R}^{2m}} \exp\left(\gamma(\theta_0)'D(\theta_0)x\right) \exp\left(x'\Omega(\Theta_0)^{-1}x\right) dx \\
&\quad (\text{where } c \text{ is a non-negative constant}) \\
&= c_1 \exp\left(-\frac{1}{2}\gamma(\theta_0)'\gamma(\theta_0)\right) \exp\left(\frac{1}{4}\gamma(\theta_0)'\gamma(\theta_0)\right) \\
&\quad (\text{by definition of the moment generating function of a multivariate normal}).
\end{aligned}$$

Since the boundary conditional likelihood for the GMM limiting experiment is given by

$$f_{\text{Bd}}(m(\theta_0) \mid d(\theta_0)) = c_2 \exp\left(-\frac{1}{2}m(\theta_0)'m(\theta_0)\right),$$

the ECS test rejects the null hypothesis if  $m(\theta_0)'m(\theta_0) > c$

### A.3.10 Proof of Result 4

The integrated likelihood

$$\begin{aligned}
f_1^*(\gamma(\theta_0)) &= c \exp\left(-\frac{1}{2}\gamma(\theta_0)'\gamma(\theta_0)\right) \\
&\quad \int_{\mathbb{R}^{n+1}} \exp\left(\gamma(\theta_0)'D(\theta_0)C^*(\theta^*, \delta)\right) \exp\left(-\frac{1}{2}C^*(\theta^*, \delta)'\Omega(\theta_0)^{-1}C^*(\theta^*, \delta)\right) \\
&\quad p_1(\theta^*, \delta) d\theta^* d\delta \\
&\quad (\text{where } c \text{ is a non-negative constant})
\end{aligned}$$

Since the boundary conditional likelihood for the weakly identified GMM model is given by

$$c_2 \exp \left( -\frac{1}{2} m(\theta_0)' m(\theta_0) \right).$$

Therefore, the ECS test statistic equals  $\frac{c}{c_2} \exp \left( -\frac{1}{2} d(\theta_0)' d(\theta_0) \right)$  times

$$\int_{\mathbb{R}^{n+1}} \exp \left( \gamma(\theta_0)' D(\theta_0) C^*(\theta^*, \delta) \right) \exp \left( -\frac{1}{2} C^*(\theta^*, \delta)' \Omega(\theta_0)^{-1} C^*(\theta^*, \delta) \right) p_1(\theta^*, \delta) d\theta^* d\delta$$

which is well defined by assumption GMM1 and also separately continuous in  $m(\theta_0)$  and  $d(\theta_0)$ . Since R2 holds, then Result 4 follows.

### A.3.11 Proof of Result 5

Let  $\xi = (S', T_1')'$  and

$$A \equiv \begin{pmatrix} C_0' / (C_0' \Omega C_0')^{1/2} \\ (C_0^{\perp'} \Omega^{-1} C_0^{\perp})^{-1/2} C_0^{\perp'} \Omega^{-1} \end{pmatrix}.$$

By construction

$$H_1 a = \begin{pmatrix} m_1 \\ \vdots \\ m_r \end{pmatrix}.$$

The integrated likelihood

$$\begin{aligned}
f_1^*(\xi) &= c \exp\left(-\frac{1}{2}\xi'\xi\right) \\
&\quad \int_{\mathbb{R}^r} \exp\left(\xi'Am\right) \exp\left(-\frac{1}{2}m'A'Am\right) \\
&\quad p_1(m)dm \\
&\quad (\text{where } c \text{ is a non-negative constant}) \\
&= c \exp\left(-\frac{1}{2}\xi'\xi\right) \int_{\mathbb{R}^r} \exp\left(\xi'Am\right) \exp\left(-m'\Omega^{-1}m\right)dm \\
&= c \exp\left(-\frac{1}{2}(S'S + T_1'T_1)\right) \exp\left(\frac{1}{4}(S'S + T_1'T_1)\right)
\end{aligned}$$

Since the boundary conditional likelihood for SVAR model is given by

$$c_2 \exp\left(-\frac{1}{2}S'S\right).$$

Therefore, the ECS test statistic equals

$$\frac{c}{c_2} \exp\left(-\frac{1}{2}T_1'T_1\right) \exp\left(\frac{1}{4}(S'S + T_1'T_1)\right)$$

Since R2 holds (see the argument in Result 1), then Result 5 follows.

### A.3.12 Proof of Result 6

The distributional assumption SVAR yields a limiting experiment as defined by Müller (2011):

$$\left(\hat{A} \otimes \hat{Q}^{-1/2}\right)(1/\sqrt{T}) \sum_{t=1}^T \eta_t \otimes Z_t \xrightarrow{d} \mathcal{N}_{rk}\left(AH_1 \otimes Q^{-1/2}a, \mathbb{I}_r \otimes \mathbb{I}_k\right) \quad (\text{A.39})$$

where

$$A \equiv \begin{pmatrix} C'_0/(C'_0\Sigma C'_0)^{1/2} \\ (C_0^{\perp'}\Sigma^{-1}C_0^{\perp})^{-1/2}C_0^{\perp'}\Sigma^{-1} \end{pmatrix}, \quad AH_1 \in \mathbb{R}^r, \quad Q^{-1/2}a \in \mathbb{R}^k$$

Note that:

$$AH_1 = \begin{pmatrix} C'_0H_1/(C'_0\Sigma C'_0)^{1/2} \\ (C_0^{\perp'}\Sigma^{-1}C_0^{\perp})^{-1/2}C_0^{\perp'}\Sigma^{-1}H_1 \end{pmatrix} = \begin{pmatrix} (\kappa - \kappa_0)/(C'_0\Sigma C'_0)^{1/2} \\ (C_0^{\perp'}\Sigma^{-1}C_0^{\perp})^{-1/2}C_0^{\perp'}\Sigma^{-1}H_1 \end{pmatrix} \quad (\text{A.40})$$

Therefore, the null hypothesis  $\kappa - \kappa_0$  holds if and only if the first element of the column vector  $AH_1$  equals zero. The limiting experiment (A.39) admits the following re-parameterization. Let

$$\phi \equiv AH_1/||AH_1||, \quad \omega \equiv Q^{-1/2}a/||Q^{-1/2}a||, \quad \rho \equiv ||AH_1|| ||Q^{-1/2}a||, \quad (\text{A.41})$$

$\phi$  is an element of the  $r - 1$  sphere,  $\mathcal{S}^{r-1} \equiv \{x \in \mathbb{R}^r : ||x|| = 1\}$ ;  $\omega$  is an element of  $\mathcal{S}^{k-1}$  and  $r$  is a non-negative scalar.

**Remark 18.** The normalization  $h_{11} = 1$  imposes a restriction on  $\phi$ . By construction, the full rank matrix  $A$  satisfies  $A\Sigma A' = \mathbb{I}_r$ , which implies  $\Sigma A'A = \mathbb{I}_r$ . Consequently,  $e_1\Sigma A'AH_1 = 1$  where  $e_1 = (1, 0, 0, \dots, 0) \in \mathbb{R}^r$ . Since  $e_1\Sigma A'$  defines a hyperplane in  $\mathbb{R}^r$ ,  $\phi$  is restricted to be an element of the positive half-space associated to that hyperplane, as  $e_1\Sigma A'\phi = 1/||AH_1|| \geq 0$ . We use  $\mathcal{S}_R^{r-1}$  to denote the intersection of the half-space and  $\mathcal{S}^{r-1}$ .

A canonical description of the SVAR testing problem is given by the following model. The sample space is  $R^r$ , with a typical element denoted by  $(S', T'_1, T'_2 \dots T'_{r-1})'$ . The parameter space is given by  $\mathbb{R}_+ \times \mathcal{S}_R^{r-1} \times \mathcal{S}^{k-1}$ , with typical element  $(\rho, \phi, \omega)$ . The statistical model is given by:

$$(S', T'_1, T'_2 \dots T'_{r-1})' \sim \mathcal{N}_{rk} \left( \phi \otimes \rho\omega, \mathbb{I}_r \otimes \mathbb{I}_k \right)$$

and the hypothesis of interest is

$$\mathbf{H}_0 : \phi_1 = 0 \quad vs. \quad \mathbf{H}_1 : \phi_1 \neq 0$$

.

Note that  $\phi_1 = 0$  implies

$$\begin{aligned} f(S', T'_1, \dots T'_{r-1}; \rho, \phi_0, \omega) &= (2\pi)^{-k/2} \exp \left( -\frac{1}{2} S' S \right) (2\pi)^{-(r-1)k/2} \\ &\quad \Pi_{s=1}^{r-1} \exp \left( -\frac{1}{2} (T_s - \phi_{s+1} \rho\omega)' (T_s - \phi_{s+1} \rho\omega) \right) \end{aligned}$$

Hence, boundary sufficiency is verified with  $g(S', T'_1, \dots T'_{r-1}) = (2\pi)^{-k/2} \exp \left( -\frac{1}{2} S' S \right)$ .

We now derive the ECS test associated with the following independent priors. Let  $\lambda_{\mathcal{S}^{r-1}}$  denote the surface measure of the  $(r-1)$  sphere  $\mathcal{S}^{r-1}$  (see Stroock (1999), p. 83)

$$\rho\omega \sim \mathcal{N}_k \left( \mathbf{0}, \lambda^2 \mathbb{I}_k \right) \tag{A.42}$$

and

$$\phi \sim \mathcal{U}(\mathcal{S}_R^{r-1}) \tag{A.43}$$

where  $\mathcal{U}(\mathcal{S}_R^{r-1})$  denotes the uniform measure on the restricted  $(r-1)$  sphere; this is, for any measurable subset  $\mathcal{A}$  of  $\mathcal{S}^{r-1}$ ,  $\mathbb{P}(\mathcal{A}) = \lambda_{\mathcal{S}^{r-1}}(\mathcal{S}^{k-1})^{-1} \int_{\mathcal{A}} \lambda_{\mathcal{S}^{r-1}}(d\phi)$ . We derive the integrated likelihood in three parts. In Part I we separate the likelihood  $f(S', T'_1, \dots T'_{r-1}; \rho, \phi, \omega)$  into three different components. In Part II we compute the partially integrated likelihood with respect to prior on  $\rho\omega$ . In Part III we present the

ECS Test.

**PART I:** Let  $T = (T'_1, T'_2, \dots, T'_{r-1})'$

$$\begin{aligned}
& -2 \ln f(S', T'_1, \dots, T'_{r-1}; \rho, \phi, \omega) - \text{cons} \\
&= \begin{pmatrix} S \\ T \end{pmatrix}' \begin{pmatrix} S \\ T \end{pmatrix} - \phi \otimes \rho \omega \\
&= \begin{pmatrix} S \\ T \end{pmatrix}' \begin{pmatrix} S \\ T \end{pmatrix} - 2 \begin{pmatrix} S \\ T \end{pmatrix}' \phi \otimes \rho \omega + (\rho \omega)'(\rho \omega) \\
&= S'S + T'T - 2\phi'(S, T_1, \dots, T_{r-1})'\rho\omega + (\rho\omega)'(\rho\omega)
\end{aligned}$$

**PART II:** Let  $M = [S, T_1, \dots, T_{r-1}][S, T_1, \dots, T_{r-1}]'$ . Note that:

$$\begin{aligned}
& \int_{\mathbb{R}^k} f(S', T'_1, \dots, T'_{r-1}; \rho, x) \exp\left(-\frac{1}{2\lambda^2}x'x\right) dx \\
&= \exp\left(-\frac{1}{2}\text{tr}(Q)\right) \int_{\mathbb{R}^k} \exp\left(\phi'(S, T_1, \dots, T_{r-1})'\rho\omega\right) \exp\left(-\frac{1}{2b^2}\right) dx \\
&\quad \text{(where } b^2 \equiv \lambda^2/(1 + \lambda^2)) \\
&= a_1(\pi, k, b^2) \exp\left(-\frac{1}{2}\text{tr}(M)\right) \exp\left(\frac{b^2}{2}\phi'M\phi\right)
\end{aligned}$$

where the last inequality follows from the definition of the moment generating function for a multivariate normal random variable.

**PART III:** Therefore, the ECS rejects if:

$$\int_{S_R^{r-1}} \exp\left(\frac{\lambda^2}{2(1 + \lambda^2)}\phi'M\phi\right) d\lambda_{S^{r-1}}(d\phi) > c(T; \lambda^2) \quad (\text{A.44})$$



# Appendix B

## Supplementary Material

### B.1 Supplementary Material for Chapter 2

#### B.1.1 Numerical Accuracy of Patnaik Methodology

We now compare two slightly different Patnaik (1949) methodologies the exact probabilities for the cases considered in Imhof (1961). While the original Patnaik approximation uses a central  $\chi^2$  distribution the modified methodology used in this paper uses a noncentral  $\chi^2$ . The advantage of using noncentral  $\chi^2$  distribution is that the approximation is exact in the conditionally homoskedastic case.

Table B.1 shows that both the central and the noncentral Patnaik methodologies are highly accurate, especially in the tails of the distributions considered. When the exact probability is  $\leq 15\%$ , the absolute error for both methodologies is at most 0.70% for all quadratic forms considered.

**Table B.1:** *Numerical Accuracy of Patnaik Methodology*

Quadratic Form	x	$P(Q > x)$			Absolute Error	
		Exact	Central	Noncentral	Central	Noncentral
$Q_1 = 0.6\chi_1^2 + 0.3\chi_1^2 + 0.1\chi_1^2$	0.1	94.58	91.85	91.85	2.73	2.73
	0.7	50.64	50.79	50.79	0.15	0.15
	2	12.40	13.10	13.10	0.70	0.70
$Q_2 = 0.6\chi_2^2 + 0.3\chi_2^2 + 0.1\chi_2^2$	0.2	99.36	98.68	98.68	0.68	0.68
	2	39.98	40.98	40.98	1.00	1.00
	6	1.61	1.45	1.45	0.16	0.16
$Q_3 = 0.6\chi_6^2 + 0.3\chi_4^2 + 0.1\chi_2^2$	1	99.73	99.61	99.61	0.12	0.12
	5	41.96	44.00	44.00	2.04	2.04
	12	0.87	0.80	0.80	0.07	0.07
$Q_4 = 0.6\chi_2^2 + 0.3\chi_4^2 + 0.1\chi_6^2$	1	96.66	95.22	95.22	1.44	1.44
	3	41.96	43.30	43.30	1.34	1.34
	8	0.87	0.66	0.66	0.21	0.21
$Q_5 = 0.7\chi_{6;6}^2 + 0.3\chi_{2;2}^2$	2	99.39	99.54	99.29	0.15	0.10
	10	40.87	40.46	41.09	0.41	0.22
	20	2.21	2.30	2.16	0.09	0.05
$Q_6 = 0.7\chi_{1;6}^2 + 0.3\chi_{1;2}^2$	1	95.49	97.19	94.96	1.70	0.53
	6	40.76	39.48	41.12	1.28	0.36
	15	2.23	2.46	2.16	0.23	0.07
$\frac{1}{3}Q_3 + \frac{2}{3}Q_4$	1.5	98.91	98.42	98.42	0.49	0.49
	4	34.53	35.52	35.52	0.99	0.99
	7	1.54	1.31	1.31	0.23	0.23
$\frac{1}{2}Q_5 + \frac{1}{2}Q_6$	3.5	95.63	96.05	95.47	0.42	0.16
	8	41.52	41.01	41.71	0.51	0.19
	13	4.62	4.74	4.58	0.12	0.04
$\frac{1}{4}(Q_3 + Q_4 + Q_5 + Q_6)$	3	98.42	98.37	98.22	0.05	0.20
	6	42.64	42.70	42.99	0.06	0.35
	10	1.17	1.16	1.09	0.01	0.08

NOTE:  $P(Q > x)$  in percent, where  $Q = \sum_{r=1}^m q_r \chi_{h_r; \delta_r^2}^2$  is a positive semidefinite quadratic form in independent normal random variables. The  $\chi_{h_r; \delta_r^2}^2$  are independent  $\chi^2$  random variables with  $h_r$  degrees of freedom and non-centrality parameter  $\delta_r^2$ . The quadratic forms, thresholds  $x$  and the exact probabilities are as in Imhof (1961). We show probabilities for the original central chi-squared Patnaik approximation and for the noncentral chi-square Patnaik approximation. The noncentral Patnaik approximation is used throughout the paper.

### B.1.2 Monte-Carlo Method

We first replace  $\mathbf{W}_2$  by the diagonal matrix of its eigenvalues and normalize the trace of  $\mathbf{W}_2$  to 1, which leaves the critical value  $c(\alpha, \mathbf{W}_2, x)$  unchanged. Our simulation routine takes as inputs the maximal asymptotic size of the test  $\alpha$ , the eigenvalues of the matrix  $\mathbf{W}_2$ , the threshold  $x$  and computes a Monte Carlo critical values  $c^m(\alpha, \mathbf{W}_2, x)$ .

Draw  $N$  independent multivariate normal random variables  $\mathbf{z}_v \sim N(\mathbf{0}, \mathbb{I}_K)$ . For a given  $\mathbf{C}$  and  $\mathbf{W}_2$  we use these normal draws to compute  $N$  draws from the distribution  $\gamma'_2 \gamma_2 / \text{tr} \mathbf{W}_2$ , where we set the default to  $N = 40,000$ . We then compute  $F_{\mathbf{C}, \mathbf{W}_2}^{m-1}(\alpha)$  as the sample upper  $\alpha$ -point from these  $N$  draws.

$c(\alpha, \mathbf{W}_2, x)$  is defined as the supremum of  $F_{\mathbf{C}, \mathbf{W}_2}^{-1}(\alpha)$  over the set

$$\Lambda = \left\{ \mathbf{C} \mid \frac{\mathbf{C}' \mathbf{C}}{\text{tr}(\mathbf{W}_2)} \leq x \right\}.$$

We construct a finite Monte Carlo analogue  $\Lambda^m$  with  $10 \times L$  elements with a default value of  $L = 50$ .

We draw  $\boldsymbol{\lambda}_i, i = 1, 2, \dots, L$  iid from a multivariate uniform distribution on  $[0, 1]^K$ . Then replace

$$\boldsymbol{\lambda}_1 = [0, \dots, 0, 1] \tag{B.1}$$

$$\boldsymbol{\lambda}_2 = [1, 1, \dots, 1] \text{ if } K \geq 2$$

$$\boldsymbol{\lambda}_3 = [0, \dots, 0, 1, 1] \text{ if } K \geq 3$$

$$\boldsymbol{\lambda}_4 = [0, \dots, 0, 1, 1, 1] \text{ if } K \geq 4$$

We then use  $\Lambda^m = \left\{ \mathbf{W}_2^{1/2} \boldsymbol{\lambda}_i \times \sqrt{t} / \sqrt{\boldsymbol{\lambda}_i' \mathbf{W}_2 \boldsymbol{\lambda}_i}, i = 1, 2, \dots, L; t = x - 9, \dots, x - 1, x \right\}$ .

The Monte-Carlo critical value is then given by

$$c^m(\alpha, \mathbf{W}_2, x) = \max \left\{ F_{\mathbf{C}, \mathbf{W}_2}^{m-1}(\alpha) \mid \mathbf{C} \in \Lambda^m \right\}.$$

### B.1.3 Levels and Sizes of Patnaik and Monte Carlo Critical Values

We compute Monte Carlo critical values  $c^m(5\%, \mathbf{W}_2, 10)$  and Patnaik critical values  $c^P(5\%, \mathbf{W}_2, 10)$  for 400 randomly drawn matrices  $\mathbf{W}_2$  of size  $K = 1, 2, 3, 4, 5$  and compare the size distortions of  $c^P$  and  $c^m$ .

We can assume wlog that  $\mathbf{W}_2$  is diagonal. For each of  $K = 1, 2, 3, 4$  we draw 100 vectors of eigenvalues  $eig(\mathbf{W}_2)$  iid from a uniform distribution on  $[0, 1]$ . We then replace the first vector of eigenvalues by  $[1, 0, \dots, 0]$ , the second one by  $[1, 1]$ , and the third and fourth ones by  $[1, 1, 0, \dots, 0]$  and  $[1, 1, 1, 0]$  provided  $K$  is large enough. We normalize the trace of  $\mathbf{W}_2$  to equal one. We denote the resulting set of diagonal matrices by  $\mathcal{W}_2$ .

We obtain  $c^P(5\%, \mathbf{W}_2, 10)$  and  $c^m(5\%, \mathbf{W}_2, 10)$  for every  $\mathbf{W}_2$  in  $\mathcal{W}_2$ . In computing the Monte Carlo critical values we use  $N = 40000$  draws and  $L = 50$  for the number of directions of  $\mathbf{C}$ . We conduct robustness checks with  $L = 1$  and  $L = 100$ .

The supporting web site for Ruud (2000) provides a MATLAB transcription of Imhof (1961)'s algorithm to compute  $F_{\mathbf{C}, \mathbf{W}_2}(x)$  for a given  $\mathbf{C}$  and  $\mathbf{W}_2$ . This allows us to compute the actual sizes  $F_{\mathbf{C}, \mathbf{W}_2}(c^P(5\%, \mathbf{W}_2, 10))$  and  $F_{\mathbf{C}, \mathbf{W}_2}(c^m(5\%, \mathbf{W}_2, 10))$  at an accuracy level of 0.01% for any  $\mathbf{C}$ ,  $\mathbf{W}_2$ .

We then compute the maximal sizes

$$maxsize^m(\mathbf{W}_2) = \max_{\mathbf{C} \in \Lambda^m} F_{\mathbf{C}, \mathbf{W}_2}(c^m(5\%, \mathbf{W}_2, 10))$$

and

$$maxsize^P(\mathbf{W}_2) = \max_{\mathbf{C} \in \Lambda^m} F_{\mathbf{C}, \mathbf{W}_2}(c^P(5\%, \mathbf{W}_2, 10)).$$

We find that  $|c^P(5\%, \mathbf{W}_2, 10) - c^m(5\%, \mathbf{W}_2, 10)|/c^m(5\%, \mathbf{W}_2, 10) \leq 4.4\%$  for all  $\mathbf{W}_2 \in \mathcal{W}_2$ . Moreover,

$$4.77\% \leq \maxsize^m(\mathbf{W}_2) \leq 5.26\% \quad \forall \mathbf{W}_2 \in \mathcal{W}_2$$

$$5.00\% \leq \maxsize^P(\mathbf{W}_2) \leq 5.02\% \quad \forall \mathbf{W}_2 \in \mathcal{W}_2$$

### B.3.1. Robustness of Size Distortions

One might be concerned that we find artificially small size distortions because we replace the set  $\mathbf{\Lambda}$  by a finite set  $\mathbf{\Lambda}^m$ . We therefore repeat our calculations for a much smaller set  $\mathbf{\Lambda}^{m,small}$  with  $L = 1$  and one much larger set  $\mathbf{\Lambda}^{m,large}$  with  $L = 100$  and find that the size distortions of both methodologies are robust. When we use  $\mathbf{\Lambda}^{m,small}$  we find that

$$4.77\% \leq \maxsize^m(\mathbf{W}_2) \leq 5.29\% \quad \forall \mathbf{W}_2 \in \mathcal{W}_2 \text{ and}$$

$$5.00\% \leq \maxsize^P(\mathbf{W}_2) \leq 5.02\% \quad \forall \mathbf{W}_2 \in \mathcal{W}_2.$$

When we use  $\mathbf{\Lambda}^{m,large}$  we find that

$$4.77\% \leq \maxsize^m(\mathbf{W}_2) \leq 5.26\% \quad \forall \mathbf{W}_2 \in \mathcal{W}_2 \text{ and}$$

$$5.00\% \leq \maxsize^P(\mathbf{W}_2) \leq 5.02\% \quad \forall \mathbf{W}_2 \in \mathcal{W}_2.$$

### B.1.4 Numerical Implementation of $B_{TSLs}(\mathbf{W}, \mathbf{\Omega})$ and $B_{LIML}(\mathbf{W}, \mathbf{\Omega})$

$$B_{TSLs}(\mathbf{W}, \mathbf{\Omega}) = \sup_{\beta \in \mathbb{R}, \mathbf{C}_0 \in S^{K-1}} \left| \frac{N_{TSLs}(\beta, \mathbf{C}, \mathbf{W}, \mathbf{\Omega})}{\mu^{-2} BM(\beta, \mathbf{W})} \right| \quad (\text{B.2})$$

$$= \sup_{\beta \in \mathbb{R}, \mathbf{C}_0 \in S^{K-1}} \left| \frac{tr \mathbf{S}_{12}}{\sqrt{tr \mathbf{S}_2 tr \mathbf{S}_1}} \left[ 1 - 2 \frac{\mathbf{C}_0' \mathbf{S}_{12} \mathbf{C}_0}{tr \mathbf{S}_{12}} \right] \right| \quad (\text{B.3})$$

Now we use that for any given  $\beta \in \mathbb{R}$  the maximum and minimum of  $\mathbf{C}_0' \mathbf{S}_{12} \mathbf{C}_0$  are  $\maxeval(\frac{1}{2} \mathbf{S}_{12} + \frac{1}{2} \mathbf{S}'_{12})$  and  $\mineval(\frac{1}{2} \mathbf{S}_{12} + \frac{1}{2} \mathbf{S}'_{12})$ . Denote  $\mathbf{S}_{12}^{sym} = \frac{1}{2} \mathbf{S}_{12} + \frac{1}{2} \mathbf{S}'_{12}$ . Then

$$B_{TSLs}(\mathbf{W}, \mathbf{\Omega}) = \sup_{\beta \in \mathbb{R}} \frac{\max(|tr \mathbf{S}_{12} - 2 \mineval(\mathbf{S}_{12}^{sym})|, |tr \mathbf{S}_{12} - 2 \maxeval(\mathbf{S}_{12}^{sym})|)}{\sqrt{tr \mathbf{S}_2 tr \mathbf{S}_1}} \quad (\text{B.4})$$

The function defined on the real line  $g_{TSLs}(\beta) = \frac{\max(|tr\mathbf{S}_{12} - 2mineval(\mathbf{S}_{12}^{sym})|, |tr\mathbf{S}_{12} - 2maxeval(\mathbf{S}_{12}^{sym})|)}{\sqrt{tr\mathbf{S}_2 tr\mathbf{S}_1}}$  converges to  $1 - 2\frac{mineval(\mathbf{W}_2)}{tr\mathbf{W}_2}$  as  $\beta \rightarrow \pm\infty$ .

The empirical researcher can specify  $\epsilon > 0$ , the desired fractional error relative to  $\lim_{\beta \rightarrow \pm\infty} g_{TSLs}(\beta)$ , and the number of starting points *points* for numerical maximization routines. The defaults are set to  $\epsilon = 0.001$  and *points* = 10000. The program then finds  $\beta_{range}$  such that  $\left| \frac{f_{TSLs}(\pm\beta_{range})}{\lim_{\beta \rightarrow \pm\infty} g_{TSLs}(\beta)} - 1 \right| \leq \epsilon$ . We maximize  $g_{TSLs}$  using the MATLAB routine `fminsearch` using *points* equally spaced starting points in  $[-\beta_{range}, \beta_{range}]$ . We also maximize  $g_{TSLs}$  over the range  $[-\beta_{range}, \beta_{range}]$  using the MATLAB routine `fminbnd`. Since each of these methodologies might only yield local maxima, we take the maximum over the local maxima to obtain  $B_{TSLs}(\mathbf{W}, \mathbf{\Omega})$ .

The numerical computation for  $B_{LIML}(\mathbf{W}, \mathbf{\Omega})$  is analogous with

$$g_{LIML}(\beta) = \max \left( \left| \frac{tr\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2} tr\mathbf{S}_1 - maxeval\mathbf{M}_B}{\sqrt{tr\mathbf{S}_1} \sqrt{tr\mathbf{S}_2}} \right|, \left| \frac{tr\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2} tr\mathbf{S}_1 - mineval\mathbf{M}_B}{\sqrt{tr\mathbf{S}_1} \sqrt{tr\mathbf{S}_2}} \right| \right) \quad (\text{B.5})$$

where  $\mathbf{M}_B = \frac{1}{2}(2\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2} \mathbf{S}_1) + \frac{1}{2}(2\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2} \mathbf{S}_1)'$  and

$$g_{LIML}(\beta) \rightarrow \frac{maxeval\mathbf{W}_2}{tr\mathbf{W}_2} \text{ as } \beta \rightarrow \pm\infty \quad (\text{B.6})$$

### B.1.5 Comparing Robust Critical Values to Stock and Yogo(05)

We now compare the critical values for our testing procedure to those in Stock and Yogo (2005). Assume for now that the errors are conditionally homoskedastic and serially uncorrelated, so that that  $W = \mathbf{\Omega} \otimes \mathbb{I}_K$ . We then obtain  $B_{TSLs}(\mathbf{\Omega} \otimes \mathbb{I}_K, \mathbf{\Omega}) = 1 - 2/K$  and  $B_{LIML}(\mathbf{\Omega} \otimes \mathbb{I}_K, \mathbf{\Omega}) = 1/K$ . We consider  $\alpha = 5\%$  and  $\tau = 10\%$ . We compare to the Stock and Yogo (2005) critical values for the null hypothesis that the asymptotic estimator bias exceeds 10% of the asymptotic OLS bias with size 5%.

Our generalized and simplified critical values differ from those values proposed by

Stock and Yogo (2005) for the TSLS bias even when first- and second-stage errors are perfectly conditionally homoskedastic and serially uncorrelated. We consider the Stock and Yogo (2005) 5% critical value for testing the null hypothesis that the TSLS bias exceeds 10% of the OLS bias and generalized and simplified critical values with a threshold of 10% and size 5%. Table B.2 shows that the TSLS critical values critical values are smaller than the Stock and Yogo (2005) critical values for  $K = 3, 4$  but larger than the Stock and Yogo (2005) critical values for  $K \geq 5$ . The difference between the TSLS and Stock and Yogo (2005) critical values is always less than 1. The simplified critical values exceed the Stock and Yogo (2005) critical values but the difference declines in the number of instruments  $K$ . The LIML critical values decline more rapidly with the number of instruments than either the TSLS or simplified critical values.

The TSLS critical values for our generalized procedure could differ from those in Stock and Yogo (2005) for two reasons.

First, we use the “worst-case” benchmark instead of the OLS bias. Denote the asymptotic OLS bias by  $Bias_{OLS} = \sigma_{12}/\sigma_2^2$ . For any structural error correlation  $\rho \in (-1, 1)$  there exists a structural parameter  $\beta \in \mathbb{R}$  such that

$$(\omega_{12} - \beta\omega_2^2)/(\sqrt{\omega_1^2 - \beta\omega_{12} + \beta^2\omega_2^2\omega_2}) = \rho,$$

provided that  $\mathbf{\Omega}$  is nonsingular. In the conditionally homoskedastic serially uncorrelated model, the Nagar bias is independent of the direction  $\mathbf{C}_0 \in \mathcal{S}^{K-1}$  with

$$N_{TSLS}(\beta, \mathbf{C}, \mathbf{W}, \mathbf{\Omega}) = Bias_{OLS} \left(1 - \frac{2}{K}\right) \quad (\text{B.7})$$

$$N_{LIML}(\beta, \mathbf{C}, \mathbf{W}, \mathbf{\Omega}) = -Bias_{OLS}/K \quad (\text{B.8})$$

and so

$$\sup_{\beta \in \mathbb{R}, \mathbf{C}_0 \in \mathcal{S}^{K-1}} (|N_e(\beta, \mathbf{C}, \mathbf{W}, \mathbf{\Omega})|/BM(\beta, \mathbf{W})) = |N_e(\beta, \mathbf{C}, \mathbf{W}, \mathbf{\Omega})|/Bias_{OLS} \quad (\text{B.9})$$

**Table B.2:** *TSLS and LIML critical values*

**Comparing to Stock and Yogo (2005)**

K	$c_{TSLS}$	$c_{LIML}$	$c_{Simp}$	$c_{TSLS}^{SY}$
1	23.11	23.11	23.11	N/A
2	3.00	12.17	19.29	N/A
3	8.53	8.53	17.67	9.08
4	10.23	6.70	16.72	10.27
5	11.06	5.61	16.08	10.83
6	11.52	4.87	15.62	11.12
7	11.80	4.35	15.26	11.29
8	11.99	3.96	14.97	11.39
9	12.11	3.65	14.73	11.46
10	12.19	3.41	14.53	11.49
11	12.25	3.21	14.36	11.51
12	12.29	3.04	14.21	11.52
13	12.32	2.90	14.08	11.52
14	12.33	2.78	13.96	11.52
15	12.35	2.67	13.86	11.51
16	12.35	2.58	13.77	11.50
17	12.35	2.50	13.68	11.49
18	12.35	2.42	13.60	11.48
19	12.35	2.36	13.53	11.46
20	12.35	2.30	13.46	11.45
21	12.34	2.25	13.40	11.44
22	12.34	2.20	13.35	11.42
23	12.33	2.15	13.29	11.41
24	12.32	2.11	13.24	11.40
25	12.31	2.07	13.20	11.38
26	12.31	2.04	13.15	11.37
27	12.30	2.01	13.11	11.36
28	12.29	1.98	13.07	11.34
29	12.28	1.95	13.04	11.33
30	12.27	1.92	13.00	11.32

NOTE: We show simplified, TSLS, and LIML critical values assuming conditional homoskedasticity, no serial autocorrelation, and known  $\mathbf{\Omega}$  and  $\mathbf{W}$ . The null hypothesis is that the Nagar bias exceeds 10% of the benchmark. Critical values have size 5% and are computed with the Patnaik methodology.  $c_{TSLS}^{SY}$  denotes Stock and Yogo (2005) 5% critical values of the null hypothesis that the asymptotic TSLS bias exceeds 10% of the asymptotic OLS bias.



Hence our choice of benchmark is not a source of divergence of our critical values from Stock and Yogo (2005).

Second, we define the null hypothesis in terms of the Nagar bias instead of the asymptotic estimator bias. Therefore, the only source of divergence of our TSLS critical values from Stock and Yogo (2005) critical values in Table B.2 is the Nagar bias approximation.

The simplified procedure also allows for the worst type of heteroskedasticity, serial correlation and/or clustering in the second stage, in contrast to Stock and Yogo (2005). Therefore, simplified critical values are higher than Stock and Yogo (2005) critical values, even when the first-stage errors are estimated to be perfectly conditionally homoskedastic and serially uncorrelated.

### B.1.6 Nagar Bias Approximates Asymptotic Bias

Assume that  $K \geq 3$ . We now prove that

$$\mathbb{E}[\beta_{TSLS}^*] - N_{TSLS} = O(\mu^{-4}) \quad (\text{B.10})$$

Write  $\mathbf{s}_2 = \frac{\mathbf{S}_2}{\text{tr}(\mathbf{S}_2)}$  and  $\mathbf{s}_1 = \frac{\mathbf{S}_1}{\text{tr}(\mathbf{S}_1)}$ . Define the normalized statistic

$$H \equiv \left( \mathbf{C}'_0 \mathbf{s}_1^{1/2} \mathbf{z}_u + \mathbf{z}'_v \mathbf{s}_2^{1/2} \mathbf{s}_1^{1/2} \mathbf{z}_u / \mu \right) \left( 1 + 2 \mathbf{z}'_v \mathbf{s}_2^{1/2} \mathbf{C}_0 / \mu + \mathbf{z}'_v \mathbf{s}_2 \mathbf{z}_v / \mu^2 \right).$$

Denote the numerator of  $H$  by  $a$  and the denominator by  $A$ .

Expand  $H$  as a stochastic power series expansion in  $1/\mu$  to get

$$\begin{aligned} H &= \mathbf{C}'_0 \mathbf{s}_1^{1/2} \mathbf{z}_u + \frac{1}{\mu} \left[ \mathbf{z}'_v \mathbf{s}_2^{1/2} \mathbf{s}_1^{1/2} \mathbf{z}_u - 2 \left( \mathbf{z}'_v \mathbf{s}_2^{1/2} \mathbf{C}_0 \right) \left( \mathbf{C}'_0 \mathbf{s}_1^{1/2} \mathbf{z}_u \right) \right] \\ &\quad + \frac{1}{\mu^2} \left[ \mathbf{C}'_0 \mathbf{s}_1^{1/2} \mathbf{z}_u \left( 2 \mathbf{z}'_v \mathbf{s}_2^{1/2} \mathbf{C}_0 \right)^2 - \mathbf{C}'_0 \mathbf{s}_1^{1/2} \mathbf{z}_u \mathbf{z}'_v \mathbf{s}_2 \mathbf{z}_v - 2 \mathbf{z}'_v \mathbf{s}_2^{1/2} \mathbf{C}_0 \mathbf{z}'_v \mathbf{s}_2^{1/2} \mathbf{s}_1^{1/2} \mathbf{z}_u \right] + \tilde{R} \end{aligned}$$

Denote the first three terms by  $d'$ . The expectation of the first three terms equals

$$N_{TSLs} \times \mu / \sqrt{\text{tr}(\mathbf{S}_1) / \text{tr}(\mathbf{S}_2)}.$$

Write  $\Delta A = A - 1$ . Both  $a$  and  $\mu\Delta A$  are finite polynomials in the components of  $\mathbf{z}$  with  $O(1)$  coefficients. A geometric series expansion gives  $H \times [1 - (-\Delta A)^3] = [a/(1 + \Delta A)] [1 - (-\Delta A)^3] = a \sum_{s=0}^2 (-\Delta A)^s$ . We can re-write this as  $H = a \sum_{s=0}^2 (-\Delta A)^s + (-\Delta A)^3 h$ . Now show that  $E[(-\Delta A)^3 h \mu^3] = O(1)$  as  $\mu \rightarrow \infty$ . Following the proof in Sargan (1974) write the expectation as an integral. Provided that the expectation exists, we know that

$$\begin{aligned} |E(-\Delta A)^3 h \mu^3| &\leq \frac{1}{(2\pi)^{K/2}} \int_{\mathbf{z} \in \mathbb{R}^K} |(\mu\Delta A(\mu, \mathbf{z}))^3| |h(\mu, \mathbf{z})| \exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{z}\right) d\mathbf{z} \\ &= \frac{1}{(2\pi)^{K/2}} \int_{\mathbf{z} \in \mathbb{R}^K} |(\mu\Delta A(\mu, \mathbf{z}))^3 a| \exp\left(-\frac{1}{4}\mathbf{z}'\mathbf{z}\right) \left|\frac{1}{A}\right| \exp\left(-\frac{1}{4}\mathbf{z}'\mathbf{z}\right) d\mathbf{z} \end{aligned}$$

But  $(\mu\Delta A(\mu, \mathbf{z}))^3 a$  is a polynomial in  $\mathbf{z}$  with coefficients  $O(1)$  as  $\mu \rightarrow \infty$ . Hence  $\exists$  constant  $B^*$  such that  $\forall \mathbf{z} \in \mathbb{R}^K$   $|(\mu\Delta A(\mu, \mathbf{z}))^3 a \exp^{-\frac{1}{4}\mathbf{z}'\mathbf{z}}| \leq B^*$ . Then,

$$\begin{aligned} |E(-\Delta A)^3 h \mu^3| &\leq B^* \frac{1}{(2\pi)^{K/2}} \int_{\mathbf{z} \in \mathbb{R}^K} A^{-1} \exp^{-\frac{1}{4}\mathbf{z}'\mathbf{z}} dz \\ &= 2^{K/2} B^* \mathbb{E} \left[ \frac{\mu^2/2}{(\mathbf{s}_2^{1/2} \mathbf{z}_v + \mathbf{C}_0 \mu / \sqrt{2})' (\mathbf{s}_2^{1/2} \mathbf{z}_v + \mathbf{C}_0 \mu / \sqrt{2})} \right] \quad (\text{B.11}) \end{aligned}$$

(B.11) can be bounded by the inverse moment of a non-central chi-square with noncentrality parameter proportional to  $\mu^2$  and 3 degrees of freedom, proving existence. Let  $X \sim \chi_3^2(y)$  be a non-central chi-square random variable with non-centrality parameter  $y$  and 3 degrees of freedom. Bock *et al.* (1984) show that

$$E[X^{-1}] = (\Gamma(1/2)/\sqrt{\pi}) \left(\frac{y}{2}\right)^{-\frac{1}{2}} D\left(\left(\frac{y}{2}\right)^{\frac{1}{2}}\right).$$

$\Gamma$  is the Gamma function and  $D$  is Dawson's integral. The values for  $D$  are tabulated in Abramowitz and Stegun (1964).  $D(y) = \frac{y^{-1}}{2} + O(y(-\frac{3}{2}))$  as  $y \rightarrow \infty$  proving that (B.11) is  $O(1)$  as  $\mu \rightarrow \infty$ .

By the uniqueness of the Taylor expansion the  $O(1)$ ,  $O(1/\mu)$  and  $O(1/\mu^2)$  terms

of  $d'$  and  $a \sum_{s=0}^2 (-\Delta A)^s$  must agree. Moreover, both  $d'$  and  $a \sum_{s=0}^2 (-\Delta A)^s$  are finite polynomials in normal random variables and this completes the proof.

### B.1.7 Primitive Conditions for Independent Data

We now specify a set of primitive conditions for independent (not necessarily identically distributed) data that imply Assumption HL. While Assumption HL is more general and can allow for serially autocorrelated data, this case encompasses cross-sectional heteroskedastic models with independent observations and linear panel data models with fixed effects and independent clusters. Assumption HL is implied by standard results for independent processes.

The main results of this section are summarized as follows. First, we show that a class of cross-sectional models satisfy Assumption HL and:

$$\begin{aligned}\mathbf{W} &\equiv \lim_{S \rightarrow \infty} (1/S) \sum_{s=1}^S \mathbb{E}[\mathbf{V}_s \mathbf{V}_s' \otimes \mathbf{Z}_s \mathbf{Z}_s'] \\ \boldsymbol{\Omega} &\equiv \lim_{S \rightarrow \infty} (1/S) \sum_{s=1}^S \mathbb{E}[\mathbf{V}_s \mathbf{V}_s'] \\ \widehat{\mathbf{W}} &\equiv (1/S) \sum_{s=1}^S \widehat{\mathbf{V}}_s \widehat{\mathbf{V}}_s' \otimes \mathbf{Z}_s \mathbf{Z}_s'\end{aligned}$$

where  $\widehat{\mathbf{V}}_s$  are OLS estimates of the reduced form errors of the model.

Second, we verify assumption Assumption HL in a class of linear panel data models with fixed effects and clustered data. Suppose  $\{\mathbf{Z}_s, v_{1s}, v_{2s}\}_{s=1}^S$  corresponds to the within transformation (Wooldridge (2002)) of the instrumental variables and the endogenous regressors in a linear panel data model with additive fixed effects. Assume that the data is partitioned according to  $L$  independent clusters and that the sample size ( $S \equiv L \times M$ ) grows as the number of observations per cluster ( $M$ ) stays constant and  $L$  grows to infinity. Write  $s \in S_l$  if observation  $s$  is in cluster  $l$  and allow for an

arbitrary correlation structure within clusters. In this case, Assumption HL is satisfied with

$$\begin{aligned}\mathbf{W} &\equiv \lim_{L \rightarrow \infty} (1/L) \sum_{l=1}^L \mathbb{E} \left[ (1/M) \left( \sum_{s \in S_l} (\mathbf{V}_s \otimes \mathbf{Z}_s) \right) \left( \sum_{s \in S_l} (\mathbf{V}_s \otimes \mathbf{Z}_s) \right)' \right] \\ \Omega &\equiv \lim_{L \rightarrow \infty} (1/L) \sum_{l=1}^L \left( (1/M) \sum_{s \in S_l} \mathbb{E}[\mathbf{V}_s \mathbf{V}_s'] \right) \\ \widehat{\mathbf{W}} &\equiv (1/L) \sum_{l=1}^L \mathbb{E} \left[ (1/M) \left( \sum_{s \in S_l} (\widehat{\mathbf{V}}_s \otimes \mathbf{Z}_s) \right) \left( \sum_{s \in S_l} (\widehat{\mathbf{V}}_s \otimes \mathbf{Z}_s) \right)' \right]\end{aligned}$$

where  $\widehat{\mathbf{V}}_s$  are the OLS estimates of the reduced form errors based of the model, and  $\{\mathbf{y}_s, \mathbf{Y}_s, \mathbf{Z}_s\}$  correspond to the within transformations of the variables. This is equivalent to estimating the reduced form errors with the fixed-effects estimator applied to the original panel data, clustering at the  $S_l$  level.

### Primitive Conditions for HL.1

Let  $\{\mathbf{X}_s\}_{s=1}^S$  be an independent  $\mathbb{R}^p$ -valued process. Let

1.  $\mathbb{E}[\mathbf{X}_s] = \mathbf{0}$  for  $s = 1 \dots S$
2.  $\overline{\mathbf{W}}_S \equiv (1/S) \sum_{s=1}^S \mathbb{E}[\mathbf{X}_s \mathbf{X}_s']$  is positive definite for  $S$  sufficiently large.
3.  $\mathbf{W} \equiv \lim_{S \rightarrow \infty} \overline{\mathbf{W}}_S < \infty$
4. There exists  $\delta > 0$  such that for all  $\boldsymbol{\lambda} \in \mathbb{R}^p$

$$\sum_{s=1}^S \mathbb{E} \left| \boldsymbol{\lambda} \overline{\mathbf{W}}_S^{-1/2} (\mathbf{X}_s) \right|^{2+\delta} \times S^{-(2+\delta)/2} \rightarrow 0$$

Theorem 3.1 in White (1980), pg. 729, imply that  $\sum_{s=1}^S \mathbf{X}_s / \sqrt{S} \xrightarrow{d} N_p(\mathbf{0}, \mathbf{W})$ .

**HL.1 for Cross-Sectional Heteroskedastic Models:** Let  $\{\mathbf{Z}_s, v_{1s}, v_{2s}\}_{s=1}^S$  be an independent process. Let  $\mathbf{X}_s \equiv (\mathbf{V}_s \otimes \mathbf{Z}_s)$  where  $\mathbf{V}_s = (v_{1s}, v_{2s})'$ . If  $\mathbf{X}_s$  satis-

fies assumptions 1-4 then  $\left((\mathbf{Z}'\mathbf{v}_1)'/\sqrt{S}, (\mathbf{Z}'\mathbf{v}_2)'/\sqrt{S}\right)' \xrightarrow{d} \mathcal{N}_{2K}(\mathbf{0}, \mathbf{W})$  where  $\mathbf{W} = \lim_{S \rightarrow \infty} (1/S) \sum_{s=1}^S \mathbb{E}[\mathbf{V}_s \mathbf{V}_s' \otimes \mathbf{Z}_s \mathbf{Z}_s']$

**HL.1 for Linear Panel Data Models with Fixed-Effects and Clustering:** Note that:

$$\mathbf{Z}'\mathbf{v}_i/\sqrt{S} = \frac{1}{\sqrt{L}} \sum_{l=1}^L \left( \frac{1}{\sqrt{M}} \sum_{s \in S_l} \mathbf{Z}_s v_{is} \right)$$

Let  $\mathbf{X}_l \equiv (1/\sqrt{M}) \sum_{s \in S_l} (\mathbf{V}_s \otimes \mathbf{Z}_s)$ . Since observations are independent across clusters, the process  $\{\mathbf{X}_l\}$  is independent. Therefore, the primitive conditions 1-4 yield the CLT  $\left((\mathbf{Z}'\mathbf{v}_1)'/\sqrt{S}, (\mathbf{Z}'\mathbf{v}_2)'/\sqrt{S}\right)' \xrightarrow{d} \mathcal{N}_{2K}(\mathbf{0}, \mathbf{W})$ , where the asymptotic covariance matrix  $\mathbf{W} = \lim_{L \rightarrow \infty} (1/L) \sum_{l=1}^L \mathbb{E} \left[ (1/M) \left( \sum_{s \in S_l} (\mathbf{V}_s \otimes \mathbf{Z}_s) \right) \left( \sum_{s \in S_l} (\mathbf{V}_s \otimes \mathbf{Z}_s) \right)' \right]$

### Primitive Conditions for Assumption HL.2

Let  $\{\mathcal{Z}_s\}$  be a sequence of independent random variables. Let

1.  $\mu_s = \mathbb{E}[\mathcal{Z}_s] < \infty$  for all  $s$ .
2. For some  $\delta > 0$ ,  $\lim_{S \rightarrow \infty} \sum_{s=1}^S \left( |\mathcal{Z}_s - \mu_s|^{1+\delta} \right) / s^{1+\delta} < \infty$ .
3.  $\Omega \equiv \lim_{S \rightarrow \infty} (1/S) \sum_{s=1}^S \mu_s < \infty$

By Corollary 3.9 in White (2001), pg. 35,  $(1/S) \sum_{s=1}^S (\mathcal{Z}_s - \mu_s) \xrightarrow{a.s.} 0$ .

**HL.2 for Cross-Sectional Heteroskedastic Models:** Let  $\{\mathbf{Z}_s, v_{1s}, v_{2s}\}_{s=1}^S$  be an independent process. For  $i, j = 1, 2$ , let  $\mathcal{Z}_{ijs} \equiv v_{is}v_{js}$ . If  $\{\mathcal{Z}_{ijs}\}_{s=1}^\infty$  satisfy 1-2 for all  $i, j = 1, 2$ , then  $(1/S) \sum_{s=1}^S \mathbf{V}_s \mathbf{V}_s' \xrightarrow{p} \mathbf{\Omega}$  where  $\mathbf{\Omega} = \lim_{S \rightarrow \infty} (1/S) \sum_{s=1}^S \mathbb{E}[\mathbf{V}_s \mathbf{V}_s']$

**HL.2 for for Linear Panel Data Models with Fixed-Effects and Clustering:**

Note that

$$\frac{1}{S} \mathbf{v}_i' \mathbf{v}_j = \frac{1}{L} \sum_{l=1}^L \left( \frac{1}{M} \sum_{s \in S_l} v_{is} v_{js} \right)$$

Let  $\mathcal{Z}_{ijs} = ((1/M) \sum_{s \in S_l} v_{is} v_{js})$ . Since observations are independent across clusters, it follows that  $\{\mathcal{Z}_{ijs}\}$  is an independent sequence. If the sequence satisfies 1-3, then it follows that  $(1/S) \sum_{s=1}^S \mathbf{V}_s \mathbf{V}_s' \xrightarrow{p} \mathbf{\Omega}$  where  $\mathbf{\Omega} = \lim_{L \rightarrow \infty} (1/L) \sum_{l=1}^L \left( (1/M) \sum_{s \in S_l} \mathbb{E}[\mathbf{V}_s \mathbf{V}_s'] \right)$ .

### Primitive Conditions for Assumption HL.3

Exercise 6.8, pg. 146, in White (2001) provides sufficient conditions for consistent estimation of the asymptotic variance in a multivariate linear model. Let:

1.  $\mathbf{Y}_s = \mathbf{X}_s' \boldsymbol{\beta} + \boldsymbol{\epsilon}_s$ ,  $s = 1, 2, \dots$   $\boldsymbol{\beta} \in \mathbb{R}^P$ ,  $\mathbf{Y}_s \in \mathbb{R}^N$ ,  $\mathbf{X}_s \in \mathbb{R}^{P \times N}$
2. Let  $\{\mathbf{X}_s, \boldsymbol{\epsilon}_s\}$  be an independent sequence (so that  $\phi$  is of size  $-1$ , with  $r = 1$ , see White (2001) page. 146).
3.  $\mathbb{E}[\mathbf{X}_s \boldsymbol{\epsilon}_s] = 0$  for all  $s$ .
4.  $\mathbb{E} \left| X_{spn} \epsilon_{sn} \right|^{2(1+\delta)} < \Delta < \infty$  for some  $\delta > 0$  and all  $n = 1 \dots N$ ,  $p = 1 \dots P$  and  $s$ .
5.  $\mathbf{V}_n \equiv \text{var} \left( (1/S) \sum_{s=1}^S \mathbf{X}_s \boldsymbol{\epsilon}_s \right)$  is uniformly positive definite.
6.  $\mathbb{E} \left| X_{sp_1 n} X_{sp_2 n} \right|^{2(1+\delta)} < \Delta < \infty$  for some  $\delta > 0$  and all  $n = 1 \dots N$ ,  $p_1, p_2 = 1 \dots P$  and  $s$ .
7.  $\mathbb{E}[(1/S) \sum_{s=1}^S \mathbf{X}_s \mathbf{X}_s']$  has uniformly full column rank and is uniformly positive definite.

Define

$$V_n \equiv (1/S) \sum_{s=1}^S \mathbb{E}[\mathbf{X}_s \boldsymbol{\epsilon}_s \boldsymbol{\epsilon}_s' \mathbf{X}_s']$$

and

$$\widehat{V}_n \equiv (1/S) \sum_{s=1}^S \mathbf{X}_s \widehat{\boldsymbol{\epsilon}}_s \widehat{\boldsymbol{\epsilon}}_s' \mathbf{X}_s'$$

where  $\hat{\epsilon}_s = \hat{Y}_s - \hat{X}_s' \hat{\beta}_{OLS}$ , and  $\hat{\beta}_{OLS} = (\sum_{s=1}^S \mathbf{X}_s \mathbf{X}_s')^{-1} \sum_{s=1}^S \mathbf{X}_s \mathbf{Y}_s$ . Exercise 6.8, pg. 146 (with  $\mathbf{Z}_s = \mathbf{X}_s$ ) in White (2001) implies that  $\hat{V}_n - V_n \xrightarrow{p} 0$ . In fact, the result holds for any  $\hat{\beta}$  such that  $\hat{\beta} - \beta \xrightarrow{p} 0$

**HL.3 for Cross-Sectional Heteroskedastic Models:** Let  $\{Z_s, v_{1s}, v_{2s}\}$  be an independent sequence. Let  $\mathbf{Y}_s = (\mathbf{y}'_s, \mathbf{Y}'_s)'$ ,  $\mathbf{X}_s = (\mathbb{I}_2 \otimes \mathbf{Z}_s)$ ,  $\beta = (\Gamma'_1, \Gamma'_2)'$ ,  $\epsilon_s = V_s$ ,  $\hat{\epsilon}_s \equiv \hat{Y}_s - \mathbf{X}_s' \hat{\beta}_{OLS} = \hat{V}_s$ . Note that  $\hat{\beta}_{OLS}$  corresponds to the reduced form OLS estimates for  $\Gamma$ . If 1-7 holds then

$$(1/S) \sum_{s=1}^S [\hat{\mathbf{V}}_s \hat{\mathbf{V}}_s' \otimes \mathbf{Z}_s \mathbf{Z}_s'] \xrightarrow{p} \mathbf{W} \equiv \lim_{S \rightarrow \infty} (1/S) \sum_{s=1}^S \mathbb{E} [\mathbf{V}_s \mathbf{V}_s' \otimes \mathbf{Z}_s \mathbf{Z}_s']$$

**HL.3 for Linear Panel Data Models with Fixed Effects and Clustering:** Let  $\{Z_s, v_{1s}, v_{2s}\}$  corresponds to the within transformation of the instrumental variables and the reduced form errors in a linear panel data model with fixed effects. Define  $\mathbf{Y}_l = (y_{l1}, \dots, y_{lM}, Y_{l1}, \dots, Y_{lM})'$ ,  $\mathbf{X}_l = (\mathbb{I}_2 \otimes (Z_{l1}, \dots, Z_{lM}))$ ,  $\beta = (\Gamma'_1, \Gamma'_2)'$ , and the innovations  $\epsilon_s = (v_{1l1}, \dots, v_{1lM}, v_{2l1}, \dots, v_{2lM})'$ . Since clusters are independent, the sequence  $\{\mathbf{X}_s, \epsilon_s\}$  is independent as well. In this case,  $\hat{\beta}_{OLS}$  corresponds to the fixed effects estimator for  $\Gamma$  in the reduced form model. If 1-7 holds then:

$$(1/L) \sum_{l=1}^L \left[ (1/M) \left( \sum_{s \in S_l} (\hat{\mathbf{V}}_s \otimes \mathbf{Z}_s) \right) \left( \sum_{s \in S_l} (\hat{\mathbf{V}}_s \otimes \mathbf{Z}_s) \right)' \right] \xrightarrow{p} \mathbf{W}$$

where  $\mathbf{W} \equiv \lim_{L \rightarrow \infty} (1/L) \sum_{l=1}^L \mathbb{E} \left[ (1/M) \left( \sum_{s \in S_l} (\mathbf{V}_s \otimes \mathbf{Z}_s) \right) \left( \sum_{s \in S_l} (\mathbf{V}_s \otimes \mathbf{Z}_s) \right)' \right]$

## B.1.8 Uniformity

Let  $\Gamma$  be a parameter space. We say that a testing procedure has asymptotic size  $\alpha$  in the uniform sense if

$$\lim_{n \rightarrow \infty} \sup_{\gamma \in \Gamma} \mathbb{P}_{\gamma}(T(x_n) > c) \leq \alpha$$

Equivalently, the testing procedure  $T$  has size  $\alpha$  (in a uniform sense) if under any sequence of parameter values  $\{\gamma_n\} \in \Gamma$ :

$$\limsup_{n \rightarrow \infty} \mathbb{P}_{\gamma_n}(T(x_n) > c) \leq \alpha$$

See Guggenberger (2010a), Guggenberger (2010b).

### B.1.9 Uniformity problems with pretests

Guggenberger (2010a) studies tests for a structural parameter  $\beta$  that follow a Hausman exogeneity pretest in an IV set-up. The Hausman pretest looks at a properly scaled difference between the TSLS and OLS estimators and tests the null hypothesis that their difference is zero.

Guggenberger (2010a) shows that the test  $T$  for a structural parameter  $\beta$  that follows a pretest

$$T(x) = \begin{pmatrix} \text{Reject } \beta = \beta_0 & \text{if} & \text{Pretest rejects and } W_{\text{TSLS}}(x) > 3.84 \\ \text{Reject } \beta = \beta_0 & \text{if} & \text{Pretest does not reject and } W_{\text{OLS}}(x) > 3.84 \end{pmatrix}$$

does not have asymptotic size  $\alpha$  in the uniform sense. The argument goes as follows: if the correlation between the second-stage structural error and the first-stage error is very small (local to zero), then the Hausman test will not reject the null hypothesis of exogeneity (with high probability). Therefore, the pretest will be followed by a Wald OLS statistic to test  $\beta = \beta_0$ . The problem is that for small values of the correlation parameter, the size of the Wald can be larger than 5%

Although it is not surprising that a pretest does not have the right size in the uni-



form sense (think about finite-sample size distortion of pretests), Guggenberger (2010a) shows that the problem with the Hausman test is very important: there is a sequence of parameter values (local-to-zero correlations) for which the rejection probability under the null is close to 1.

### B.1.10 Uniformity and Tests for Weak Instruments

It is true that a two-stage test that selects between a standard procedure (like the Wald) and a robust procedure (like the Anderson and Rubin test) following a test for weak instruments will in general lack uniformity. That is, a nominal  $\alpha$ -level test for weak instruments followed by a nominal  $\alpha$ -level test in a second-stage need not deliver an overall  $\alpha$ -level test under a weak instrument sequence. However, the size distortion need not be large.

To illustrate this point consider the following just-identified model with arbitrary heteroskedasticity, serial correlation and/or clustering. The argument used in this example extends straightforward to an IV model with conditional homoskedastic serially uncorrelated errors. Suppose we test the null hypothesis:

$$\mathbf{H}_0 : \beta^* = \beta_0 \quad vs. \quad \mathbf{H}_0 : \beta^* \neq \beta_0$$

For simplicity, consider the following finite sample assumption:

$$\begin{bmatrix} \sqrt{T}\widehat{\beta\pi} \\ \sqrt{T}\widehat{\pi} \end{bmatrix} \equiv \begin{bmatrix} (Z'Z/T)^{-1}Z'y/\sqrt{T} \\ (Z'Z/T)^{-1}Z'Y/\sqrt{T} \end{bmatrix} \sim \mathcal{N}_2 \left( \begin{pmatrix} \beta^*c \\ c \end{pmatrix}, \mathbf{\Omega} \right),$$

with  $\Omega$  known, so that the parameter space is  $(\beta^*, c) \in \mathbb{R} \times \mathbb{R}$ . Let

$$\begin{aligned}\widehat{\beta}_{IV} &= \widehat{\beta\pi}/\widehat{\pi} \\ F_T^e &= (\sqrt{T}\widehat{\pi})^2/\omega_2^2 \\ AR_T(\beta_0) &= (\sqrt{T}\widehat{\beta\pi} - \beta_0\sqrt{T}\widehat{\pi})^2/(1, -\beta_0)\mathbf{\Omega}(1, -\beta_0)' \\ W_T(\beta_0) &= (\sqrt{T}\widehat{\beta\pi} - \beta_0\sqrt{T}\widehat{\pi})^2/(1, -\widehat{\beta}_{IV})\mathbf{\Omega}(1, -\widehat{\beta}_{IV})'\end{aligned}$$

Consider the test that follows the pretest that uses the critical value of 23:

$$T(\sqrt{T}\widehat{\beta\pi}, \sqrt{T}\pi) = \begin{pmatrix} \text{Reject } \beta^* = \beta_0 & \text{if } F_T^e < 23 \text{ and } AR_T(\beta_0) > 3.84 \\ \text{Reject } \beta^* = \beta_0 & \text{if } F_T^e > 23 \text{ and } W_T(\beta_0) > 3.84 \end{pmatrix}$$

The finite sample size is given by:

$$\sup_{c \in \mathbb{R}} \left[ \mathbb{P}_{\beta_0, c} \left( F_T^e < 23 \cap AR_T(\beta_0) > 3.84 \right) + \mathbb{P}_{\beta_0, c} \left( F_T^e > 23 \cap W_T(\beta_0) > 3.84 \right) \right]$$

Stock and Yogo (2005) have shown that for any  $\beta_0$  in this Gaussian Model:

$$\sup_{c^2/\omega_2^2 > 6} \mathbb{P}_{\beta_0}(W_T > 3.84) \leq 10\%.$$

That is, the size distortion is smaller than 5% if the “concentration parameter”  $c^2/\omega_2^2 > 6$ . Furthermore, the maximal size distortion is decreasing in such parameter. In the just-identified case a value of  $c^2/\omega_2^2 > 10$  guarantees that the relative Nagar bias is smaller than 10%. Also, the rejection probability of the effective F (in the just-identified model) is decreasing in such parameter.

Therefore, we can bound the maximal rejection probability by considering the following

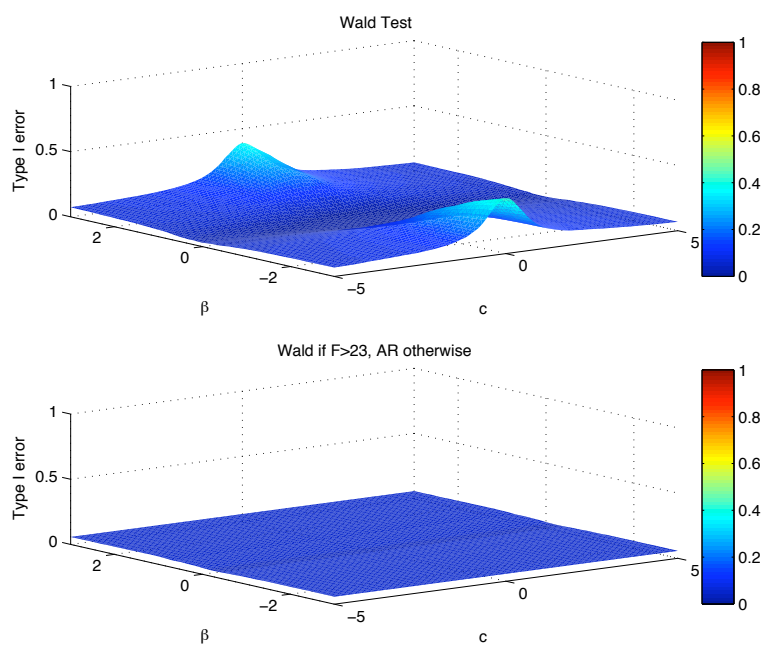
inequalities. First:

$$\begin{aligned}
& \sup_{c^2/\omega_2^2 > 10} \left[ \mathbb{P}_{\beta_0, c} \left( F_T^e < 23 \cap AR_T(\beta_0) > 3.84 \right) + \mathbb{P}_{\beta_0, c} \left( F_T^e > 23 \cap W_T(\beta_0) > 3.84 \right) \right] \\
& \leq \sup_{c^2/\omega_2^2 > 10} \left[ \min \left( \mathbb{P}_{\beta_0, c} (F_T^e < 23), 5\% \right) + \min \left( \mathbb{P}_{\beta_0, c} (F_T^e > 23), \mathbb{P}_{\beta_0, c} (W_T(\beta_0) > 3.84) \right) \right] \\
& < 15\%
\end{aligned}$$

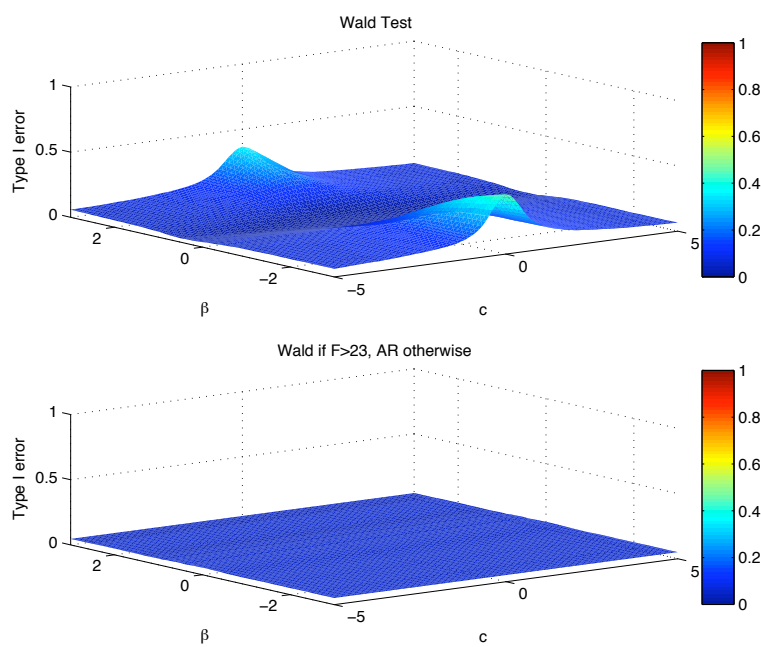
Second, for values of  $c^2/\omega_2^2$  for which the Nagar relative bias is larger than 10%:

$$\begin{aligned}
& \left[ \mathbb{P}_{c^2/\omega_2^2 < 10} \left[ \mathbb{P}_{\beta_0, c} \left( F_T^e < 23 \cap AR_T(\beta_0) > 3.84 \right) + \mathbb{P}_{\beta_0, c} \left( F_T^e > 23 \cap W_T(\beta_0) > 3.84 \right) \right] \right] \\
& \leq \sup_{c^2/\omega_2^2 < 10} \left[ \min \left( \mathbb{P}_{\beta_0, c} (F_T^e < 23), 5\% \right) + \mathbb{P}_{\beta_0, c} (F_T^e > 23) \right] = 10\%
\end{aligned}$$

Therefore, a conservative upper bound for the uniform size of the test that follows our pretest for weak instruments is 15%. The same argument holds in the conditionally homoskedastic, serially uncorrelated, over-identified IV model. To illustrate our point, we report Monte Carlo simulations with 5000 draws for the size of the two-stage test  $T$ . Without loss of generality we consider a covariance matrix  $\mathbf{\Omega}$  with unit variances and correlation parameter  $\omega_{12} = \rho \in [0, 1]$ . We report the size of the test  $T$  for  $\rho = .2, .4, .6, .8$ , a uniform grid for  $c \in [-5, 5]$  and a uniform grid for  $\beta \in [-3, 3]$ .

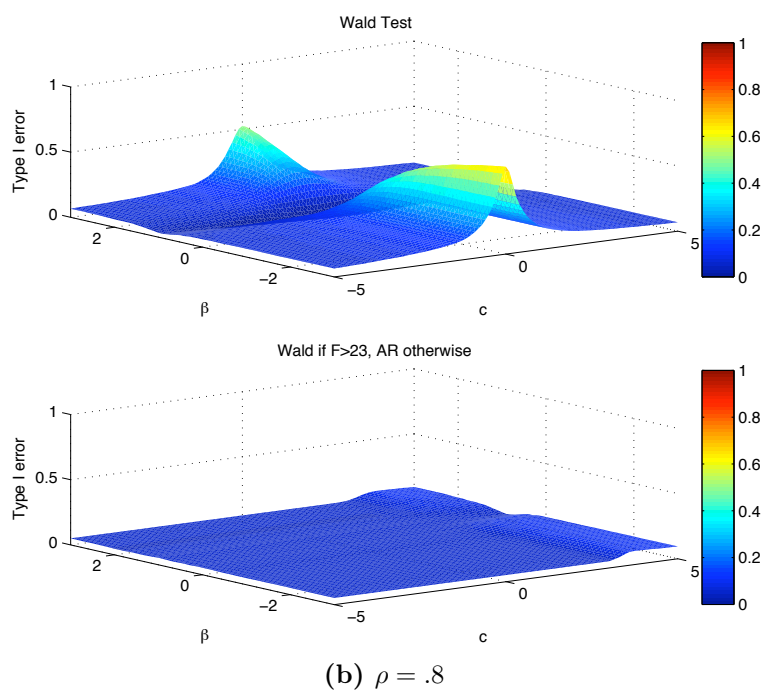
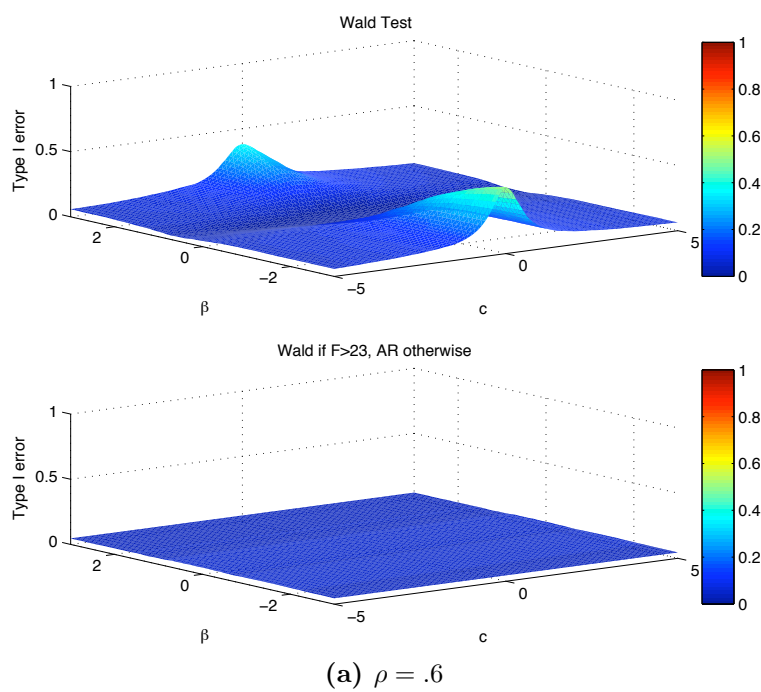


(a)  $\rho = .2$



(b)  $\rho = .4$

**Figure B.1:** *Rejection Probabilities for the Wald and T Tests ( $\alpha = 5\%$ )*



**Figure B.2:** *Rejection Probabilities for the Wald and T Tests ( $\alpha = 5\%$ ), ctd.*